Filamentous Phage Infection of *E. coli:* Mechanistic Aspects and Applications to *in vitro* Protein Evolution

---

DISSERTATION

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Alexandre Mooser

von

Jaun FR

Promotionskomitee

Prof. Dr. Andreas Plückthun (Vorsitz)

Prof. Dr. Hans Rudolf Bosshard

Prof. Dr. Raimund Dutzler

Zürich 2005

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich im Wintersemester 2005/2006 als Dissertation angenommen.

Promotionskomitee:

Prof. Dr. Andreas Plückthun (Vorsitz)

Prof. Dr. Hans Rudolf Bosshard

Prof. Dr. Raimund Dutzler

Erklärung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe im Sinne von §3 und §5 der Promotionsverordnung vom 08. Juli 2002 angefertigt. Bei der Abfassung der Dissertation wurden keine anderen als die darin angegebenen Hilfsmittel benützt.

Zürich, Dezember 2005

Alexandre Mooser

PREFACE

Bacterial conjugation, as far as we know, is a uniquely prokaryotic phenomenon, reflecting a fundamental structural distinction between prokaryotic and eukaryotic cells. Eukaryotic cells, with their genome sequestered from the plasma membrane by inclusion in intracellular organelles, have not evolved or retained mechanisms of genetic exchange that require the movement of DNA directly from one cell to another. Prokaryotic cells, lacking the constraints imposed by such organelles, have widely exploited such mechanisms of horizontal transfer. In spite of its near-ubiquity in the bacterial world, and its clear relevance to the dissemination of antibiotic resistance, conjugal DNA transfer remains a “black box”. The organism responsible for the horizontal transfer of genetic material among prokaryotes was identified more than half a century ago.

In 1940, Max Delbrück and Salvador Luria initiated a series of studies with a new organism, destined to become as important to genetic research as the garden pea and the fruit fly. This new organism was a group of viruses that attack bacterial cells and are therefore known as bacteriophage (“bacteria eater”) or simply, phage. Every known type of bacterial cells is preyed upon by its own type of bacterial virus, and many bacteria are host to many different kinds of viruses. It was later found that phage are the key players of dissemination of genetic material in the bacterial world.

In 1985, G.P. Smith proved the concept of “phage display”: fragments of the *EcoRI* endonuclease could be fused to the amino-terminal portion of a phage coat protein, to produce a chimeric protein that was packaged into phage particles. These phages could be propagated in bacteria, even though they were somewhat defective with regard to infectivity. They could be affinity-purified from a mixture of phages in which the majority did not contain the *EcoRI* insert.

The present manuscript explores the world of bacteriophages. More specifically, it aims at a better understanding of the processes governing bacterial infection by filamentous bacteriophages. A better understanding of infection mechanisms has implications for biotechnology, concerning phage display technology, as well as for medicine, concerning the association of bacterial virulence genes with mobile genetic elements like the filamentous phages ctxϕ, in the case of cholera.

Felix d’Hérelle was a French-Canadian bacteriologist. In 1910, he was studying a plague of locusts in Mexico when he noticed that they were being killed by some natural agent, which he called bacteriophage – devourer of bacteria

Acknowledgments

I would like to thank Prof. Dr. Andreas Plückthun for his constant support through these projects, his patience, encouragement to independence, and for offering me the opportunity to conduct my doctoral thesis in his research group. Prof. Dr. Hans-Rudolf Bosshard and Raimund Dutzler are kindly acknowledged for accepting to judge this thesis.

The realization of several different projects gave me the opportunity to collaborate with several people whom I am greatly indebted to, and whom I would like to thank here: Prof. Dr. Alex Wlodawer and Dr. Changsoo Chang from the National Cancer Institute, Bethesda (MD) for their fruitful collaboration on the TonB project. I also would like to address special thanks to Daniela Roethlisberger and Frédéric Pecorari for inspiring ideas and enriching discussions on the phage display project. Many thanks to Patrik Forrer for his invaluable help and assistance with methodical issues which have broadened my knowledge of biochemistry and protein chemistry. I would like to thank Prof. Dr. Hideo Iwai for his support with protein structure-related questions and for helping me interpret results of his NMR measurements.

Many thanks to all non-explicitly mentioned former and present members of the AP group, for helping me out and for creating such a pleasant lab environment, and especially to Dr. Peter Lindner for all assistance with organizational procedures.

The biggest thank goes to my parents and my family for their constant support throughout my studies.

Summary

The present work is dedicated, in its first part, to the detailed study of the structural processes governing bacterial infection by filamentous phage. In a second part, I illustrate the use of filamentous phage as a tool to improve the binding property of an antibody fragment.

The first section, entitled “Biology and structure of filamentous bacteriophages”, is intended to give the reader a general overview of the biology of filamentous phages and bacteria, focussing on those concepts that are useful for the understanding of the experiments in the subsequent chapters of this dissertation.

In the second section, two proteins which are intricately implicated in the infection process of *Escherichia coli* by filamentous bacteriophages have been investigated and, to some extent, biochemically characterized. The minor coat protein pIII has been extensively studied; its biochemical characterization led us to the conclusion that pIII is intrinsically unstable in solution. Once isolated from inclusion bodies in *E. coli*, it is proteolytically processed at its C-terminus. The contaminating protease responsible for this spontaneous effect has remained elusive. Different protein constructs were generated, and they all showed the same pattern of unspecific cleavage. The next chapter of this section is devoted to the TolQRA complex, and in particular to the inner membrane-anchored TolR, known to be critical for phage entry. This protein has proved to possess a high degree of conformational flexibility. Different C-terminal constructs of various lengths were assessed by two-dimensional NMR spectra; they all showed a non-optimal dispersion pattern, thus preventing any further detailed structural characterization. Finally, the topologically related ExbB-ExbD-TonB complex was investigated. This complex provides energy for siderophore and colicins transport across the outer membrane, but it has also been parasitized by certain phages. Using a prediction-based method to define the folded domain of TonB, a crystal structure could be obtained. Our published crystal structure of the C-terminal domain of TonB forms the basis of a critical discussion about the transport mechanism of colicins into *E. coli* (Chang et al., 2001). Our findings have been challenged by other research groups, who concluded that our dimeric structure resulted from a crystallization artefact. New results on TonB itself, as well as on outer membrane transporters contacted by the C-terminus of TonB are reviewed and critically discussed.

The third section demonstrates the usefulness of bacteriophages in a modern biotechnological application. It is illustrated by the affinity maturation of an antibody fragment directed against a mammalian membrane protein complex of immunological relevance. By combining phage display and structure-based antibody engineering methods, we tried to improve the dissociation constant of a Fab antibody fragment specific for the MHC-gp33 peptide complex. By subjecting the variable domains of the starting clone to error-prone PCR, several libraries with different mutational loads were constructed. The libraries were displayed monovalently at the

tip of the phage, and off-rate selection was applied in order to select for higher affinities. Unexpectedly, the heavy chain of the original wild-type Fab has been reselected after two rounds of phage display. By constructing a single domain antibody consisting solely on the wild-type heavy chain of the variable domain, full binding activity towards the MHC-gp33 was retained, but the affinity did not improve as expected. An interpretation of our results is given in view of recent protein evolution studies dealing with protein plasticity and their tolerance to amino acids mutations.

Literature

Chang, C., Mooser, A., Plückthun, A., and Wlodawer, A. (2001). Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. J Biol Chem 276, 27535-27540.

Zusammenfassung

Die vorliegende Arbeit ist in ihrem ersten Teil der ausführlichen Untersuchung der strukturellen Prozesse gewidmet, die die bakterielle Infektion durch filamentöse Bakteriophagen regeln. In einem zweiten Teil wird der Gebrauch der filamentösen Bakteriophagen als Werkzeug beschrieben, um die Bindungs-eigenschaft eines Antikörperfragments zu verbessern.

Der erste Abschnitt, betitelt „Biology and structure of filamentous bacteriophages", soll dem Leser einen allgemeinen Überblick über die Biologie der filamentösen Bakteriophagen und des Bakteriums geben. Er fokussiert auf einige Konzepte, die für das Verständnis des Fachstoffes in den folgenden Kapiteln nützlich sind.

Im zweiten Abschnitt wird die Untersuchung und biochemische Charakterisierung zweier Proteine beschrieben, von denen man annimmt, dass sie im Infektionsprozeß von *Escherichia coli* durch filamentöse Bakteriophagen involviert sind. Das Hüllprotein pIII ist im Detail untersucht worden; seine biochemische Charakterisierung führte uns zum Ergebnis, dass pIII in gelöster Form an sich instabil ist. Kurz nach Isolierung aus den „inclusion bodies" in *E. coli* wird das pIII an seinem C-terminalen Ende proteolytisch prozessiert. Die kontaminierende Protease, die für diesen spontanen Effekt verantwortlich ist, ist schwer bestimmbar geblieben. Unterschiedliche Proteinkonstrukte wurden erzeugt, und sie zeigten alle das gleiche Muster von unspezifischer Spaltung. Das folgende Kapitel dieses Abschnitts wird dem TolQRA Komplex und insbesondere einem seiner Mitglieder gewidmet. Das in der inneren Membran verankerte TolR, bekannt für seine ausschlaggebende Rolle in Bakteriophageninfektion, besitzt einen hohen Grad konformationeller Flexibilität. Unterschiedliche C-terminale Konstrukte von verschiedenen Längen wurden durch zweidimensionale NMR Spektren untersucht; alle zeigten eine nicht-optimale Dispersion und verhinderten jede weitere ausführliche strukturelle Charakterisierung. Schließlich wird ein Kapitel dem topologisch verwandten ExbB-ExbD-TonB Komplex gewidmet. Dieser Komplex stellt Energie für den Transport von Siderophoren und Colicinen durch die äußere Bakterienmembrane zur Verfügung, aber er wird auch durch bestimmte Phagen ausgenützt. Durch die Definition und erfolgreiche Expression und Kristallisation der gefalteten Domäne des TonB konnte dessen Kristallstruktur bestimmt werden. Unsere veröffentliche Kristallstruktur der C-terminalen Domäne von TonB bildet die Grundlage einer kritischen Diskussion über die Transportmechanismen von Colicinen in *E. coli* (Chang et al., 2001). Unsere Entdeckungen wurden von anderen Forschungsgruppen in Frage gestellt, die feststellten, dass unsere dimere Struktur aus einem Kristallisationsartefakt resultieren könnte. Neue Ergebnisse mit TonB selbst, sowie mit den Transportern der äusseren Membran, die mit dem C-Terminus von TonB in Verbindung treten, werden zusammengefasst und kritisch besprochen.

Der dritte Abschnitt zeigt den Nutzen von Bakteriophagen in einem modernen biotechnologischen Prozess. Die Anwendung der „Phage Display“ wird durch die Affinitätsmaturierung eines Antikörperfragments veranschaulicht, das gegen einen Membranproteinkomplex von immunologischen Bedeutung gerichtet ist. Indem wir „Phage Display“ und neue Antikörpermethoden kombinierten, versuchten wir, die Dissoziationskonstante eines Fab Antikörperfragments spezifisch für den MHC-gp33 Peptidkomplex zu verbessern. Die variable Domäne des Startklons wurden einer „error-prone“ PCR unterzogen, und mehrere Bibliotheken mit unterschiedlichen Mutationsraten wurden gebaut. Die Bibliotheken wurden monovalent an der Phagenspitze exprimiert, und „off-rate“ Selektion wurde angewandt, um für höhere Affinitäten zu selektionieren. Unerwartet ist die schwere Kette des ursprünglichen Wildtyp Fab nach zwei Runden „Phage Display“ wieder ausgewählt worden. Ein Antikörperfragment, das nur auf der variablen schwere Kette besteht, wurde konstruiert. Bindung an den MHC-gp33 Peptidkomplex wurde beibehalten, aber die Affinität verbesserte sich nicht wie erhofft. Unsere Ergebnisse werden im Hinblick auf neue Protein-Evolutionsstudien interpretiert, die sich mit Proteinplastizität und ihrer Toleranz gegen Aminosäureveränderungen befassen.

Literatur

Chang, C., Mooser, A., Plückthun, A., and Wlodawer, A. (2001). Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. J Biol Chem 276, 27535-27540.

Table of Contents

SECTION 1
INTRODUCTION

1.1 Biology and structure of filamentous bacteriophages

The filamentous bacteriophages (Genus *Inovirus*) constitute a group of prokaryotic viruses that contain a circular single-stranded DNA genome encased in a long protein capsid cylinder. The relative simplicity of theses viruses and the ease with which they can be manipulated have made them fruitful models to study macromolecular structure and interactions. Many different biological strains have been characterized, with DNA sequence homology ranging from very little to almost complete, but all strains have a similar virion structure and life cycle. The Ff class of the filamentous bacteriophages (f1, fd and M13) have been the most extensively studied. The present work concentrates mostly on this class of bacteriophages, which use the F conjugative pilus as a receptor and thus are specific for *Escherichia coli* containing the F plasmid. The DNA sequence of these three phages shows them to be 98% homologous. Phage particles do not form in the cytoplasm; rather they are continually extruded or secreted across the bacterial membranes as they are assembled, without causing cell lysis. A mixture of bacterial and phage-encoded components allow the single-stranded viral DNA to be replicated via a double-stranded intermediate. The result of this replicative process is a newly synthesized viral single-stranded DNA in a complex with many copies of a phage-encoded single-stranded DNA-binding protein.

All five structural proteins of the virion are synthesized as integral membrane proteins and are anchored in the inner membrane of *E. coli* prior to their incorporation into phage particles. Assembly occurs at specific sites in the bacterial envelope where the inner and the outer membrane are in close contact (Lopez and Webster, 1985). These contact sites, still a matter of controversy, can be visualized by electron microscopy and may reflect the existence of special assembly sites or exit complexes necessary for phage assembly and extrusion. Phage assembly and extrusion is a membrane-asssociated event, fueled by ATP hydrolysis, the membrane potential (Feng et al., 1997) and at least one bacterial protein, thioredoxin. During the assembly process, the viral DNA is extruded through the membrane-associated assembly site, where the phage DNA-binding protein is replaced by capsid proteins which wrap around the DNA. This process continues until the end of the DNA is reached, so there is little if any constraint on the size of the DNA packaged. The Ff bacteriophages have a single-stranded DNA molecule of about 6400 nucleotides, but foreign DNA up to 12’000 nucleotides long can be inserted into a non-essential region of the viral genome, and this feature has made the phages a powerful tool in biotechnology. By inserting foreign DNA as a separate gene, one can create a virus-derived plasmid (or phagemid)

which can become a cloning vector. When the foreign DNA is fused in-frame to the gene coding for a viral coat protein, the virion can display foreign peptides or proteins on its surface. This technology and its applications are discussed in greater detail in the third section of this work (“Application Of Phage Display”).

The Ff phage particle is a rod about 6.5 nm in diameter and 930 nm in length. It has a mass of about 16.3 MegaDalton, of which 87% is made of protein. It comprises a sheath of 2’700 identical major coat proteins, also called *gene VIII* protein (pVIII), in an helical array around a single-stranded circular DNA molecule of about 6400 nucleotides (Figure 1-1). Minor coat proteins cap each end. The packaging signal end, which is always the first to emerge from the cell, contains about 5 copies each of the 33-residue *gene VII* protein (pVII) and the 32-residue *gene IX* (pIX), necessary for efficient particle assembly. Genetic and immunological experiments trying to model the ends of the particle support the hypothesis that pVII is shielded from the environment, whereas pIX is exposed (Endemann and Model, 1995). pIX has been found to be mainly α-helical (Houbiers et al., 1999). The other end contains 3-5 copies each of the 406-residue *gene III* and 112-residue *gene VI* proteins (pIII and pVI), which are both required for particle stability and phage infectivity. Biochemical evidence suggests that pIII and pVI interact with each other in the phage particle, as they remain associated after disruption of the phage particle with detergents like sodium deoxycholate. The formation of the pIII-pVI complex is essential for correct termination of the phage assembly (Gailus et al., 1994); pIII is thought to protect pVI from rapid degradation (Rakonjac and Model, 1998). No structure of pVI has ever been reported, but its hydrophobic amino terminus is believed to be buried in the phage coat (Makowski, 1992), whereas the C-terminus might be solvent-exposed allowing the attachment of foreign proteins at this site (Jespers et al., 1995)

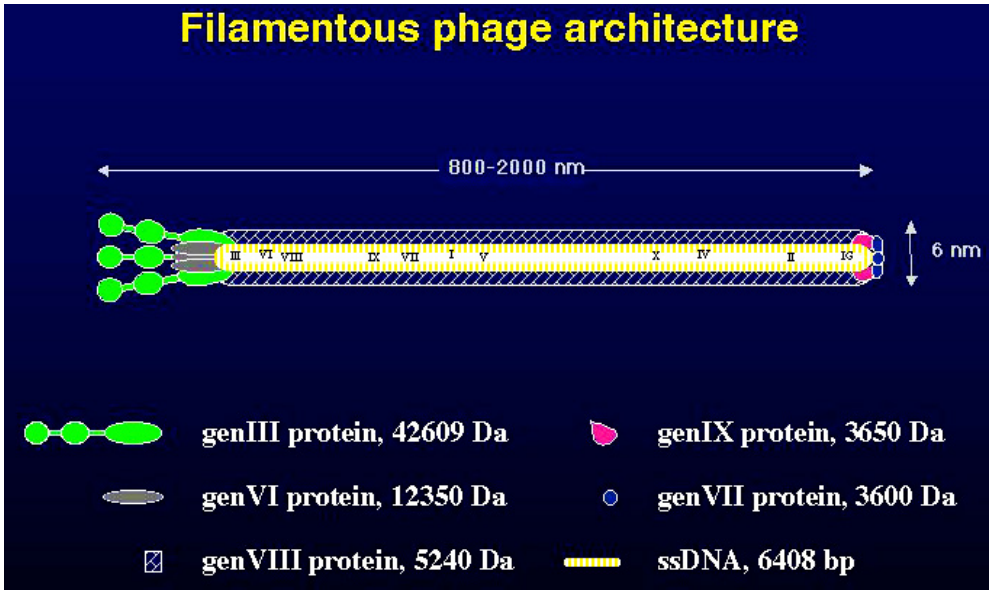


Figure 1-1. Schematic representation of a phage particle. pIII is believed to be present in three to five copies at the tip of the phage.

The pVIII portion of the virion has been extensively characterized and its uninterrupted α-helical structure solved by NMR (Marassi and Opella, 2003). Before being assembled into phage particles, the hydrophobic segment of pVIII spans the inner membrane of E. coli, with the negatively charged N-terminal segment outside in the periplasm and the positively charged hydrophilic C-terminal segment inside in the cytoplasm (Figure 1-2). In the native virion, the carboxy-terminal 10-13 residues of pVIII line the inner surface of the sheath, where they neutralize the negative charge of the DNA core (Symmons et al., 1995). The amino-terminal portion of pVIII is present on the outside of the particle.

pIII is made up of three domains, separated by glycine-rich peptide linkers (Stengele et al., 1990). It is synthesized as a precursor (424 amino acids), in which the mature protein is preceded by an 18-residue signal sequence. Deletion analysis led to the conclusion that the N-terminal domain (N1) is mainly responsible for membrane penetration, the middle domain (N2) for adsorption to the F-pilus (Deng et al., 1999) and the sole function assigned to the C-terminal domain (CT domain) is the anchoring of pIII in the phage coat. No proper function has been assigned to the glycine-rich linkers, which are believed to mainly convey flexibility to the other domains of pIII.

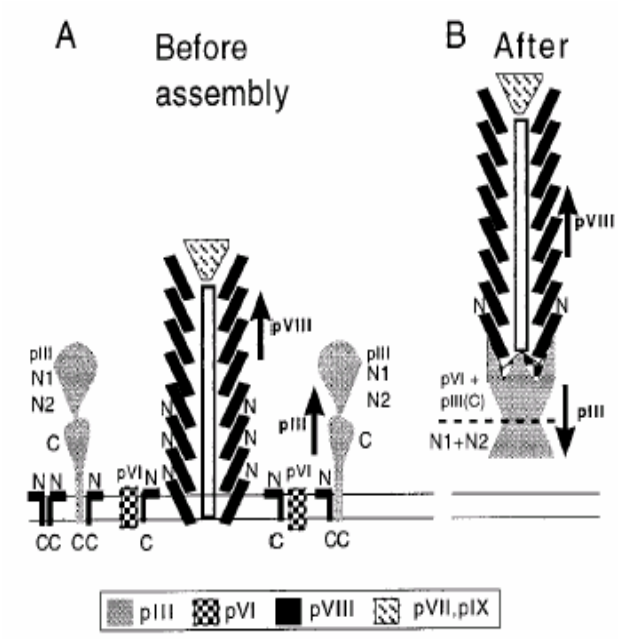


Figure 1-2. Cartoon of the termination of assembly. (a) Before termination, pIII, pVI and pVIII are integral membrane proteins. Both pVI and pIII interact with pVIII. The N terminal portions of pIII and pVIII are located in the periplasm. pIII is composed of N1, N2 and C domains. N and C depict respectively, the N and C-termini of pIII and pVIII. The C to N direction, which is the same for both proteins, is indicated by the arrows. The orientation of pVI is not known. (b) After termination, the N terminal domain of pIII points in the opposite direction from the N-terminus of pVIII. The white rod along the axis of the virion symbolizes ssDNA. The two parallel horizontal lines represent the inner membrane of *E. coli*, with the periplasm at the top and the cytoplasm at the bottom. Picture taken from Rakonjac et al. (1999).

The N1-N2 fragment of pIII forms a bilobal, horseshoe-like structure, in which the larger N2 domain partially wraps around the smaller N1 domain. The large number of exposed aliphatic residues (from N1) and threonine residues (from N2) pointing into the central channel have made this central cavity an attrative potential interaction site for the pilus (Holliger and Riechmann, 1996; Holliger et al., 1999; Lubkowski et al., 1998). Nevertheless, mutagenesis showed that the binding site is on the outer surface of N2 (Deng and Perham, 2002). N2 is in fact composed of two subdomains: a globular part, which resembles N1 in size and tertiary structure, and a hinge subdomain, which provides an extended surface for interactions between N1 and N2 (Figure 1-3). Kinetic studies have established the importance of this hinge region. Folding of domain N1 takes a few milliseconds, and provides a framework for the subsequent folding of the N2 domain, which is achieved in about five minutes. The final reaction in the folding of pIII involves the docking between N1 and N2, and its rate is controlled by the *trans* to *cis* isomerization of the Gln212-Pro213 prolyl peptide bond in the hinge subdomain of N2, which was identified as the rate-determining event of the docking reaction. The very slow domain docking reaction (about 100 minutes) constitutes a protection against unfolding for the two domains, and has major implications with regard to the function of pIII in the infection pathway (see below). From these studies, it was concluded that the proline-controlled domain docking is a late step occurring after the two domains have reached their final three-dimensional conformation (Martin and Schmid, 2003a; Martin and Schmid, 2003b; Martin and Schmid, 2003c). Under certain conditions, this portion of pIII can be visualized by electron microscopy as a knob-like structure at one end of the particle (Gray et al., 1981).

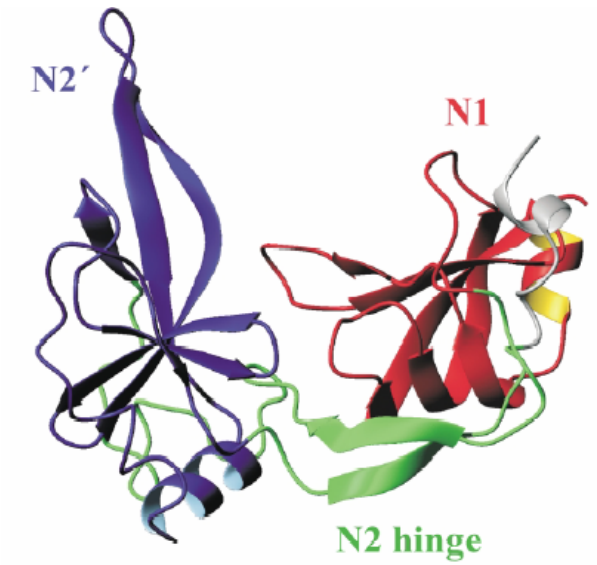


Figure 1-3. Ribbon representation of the tertiary structure of pIII: the N1 domain is in blue, N2 domain in red, and the hinge subdomain of N2 in green. The glycine-rich linker, due to its conformational lability, is not resolved. Picture taken from Martin and Schmid (2003c).

The carboxy-terminal 150 residues of pIII is required for the release of newly assembled virions from the membrane and for structural stability of the phage particle (Armstrong et al., 1981; Crissman and Smith, 1984; Stengele et al., 1990). Detailed analysis of deletion mutants revealed that the C-terminal domain of pIII contains two functionally distinct subdomains: one is involved in capping and stabilizing the assembled phage particle, and the second one is required for incorporation into the phage coat and subsequent release of the assembled virion from the host membrane (Rakonjac et al., 1999). Although filamentous phage infection does not usually kill the host cells, the infection can be lethal if assembly is impaired.

For more information about bacteriophage structure, see Marvin (1998), Marvin et al. (1994), Russel et al. (1997).

1.2 The infection process and the TolQRA pathway

The *tol-pal* gene cluster coding for seven genes is located at 17 minutes on the + strand of the *E. coli* chromosome (Webster, 1991). Currently the precise function encoded by the gene cluster is unclear, however they appear to be involved in maintaining the integrity of the cell envelope, since mutations in any of the *tol-pal* genes result in hypersensitivity to deleterious agents, release of periplasmic proteins (Lazdunski et al., 1998), formation of outer-membrane vesicles at the cell surface and induction of capsule synthesis, which results in a mucoid phenotype (Bernadac et al., 1998). The Tol-Pal gene products are organized into two complexes: one is composed of TolA, TolQ and TolR and sits in the inner membrane. The other is located in the outer membrane and comprises of TolB, Pal, Orf2 and two proteins of the outer membrane, Lpp and OmpA (Figure 1-4 and 1-5). There is no direct evidence of interactions between the inner membrane and outer membrane complexes; however, the localization of Tol proteins in contact sites between the inner and outer membranes support this hypothesis (Guihard et al., 1994). The proteins TolQ, TolR and TolA are all associated with the inner membrane and appear to interact with each other via their transmembrane (TM) domains (Lazzaroni et al., 1995). The TolA TM domain interacts with the first TM domain of TolQ and with TolR (Derouiche et al., 1995; Germon et al., 1998), while the third TM domain of TolQ interacts with the TM domain of TolR. TolR and TolA have a three-domain structure. In addition to the N-terminal anchoring region, TolA contains a large central domain with a high degree of α–helical structure and a C-terminal domain whose crystal structure has been solved (Lubkowski et al., 1999; Witty et al., 2002). The second domain contains an alanine-enriched coiled-coil region, which seems likely to play a role in communicating signals from the inner to the outer membrane of *E. coli*. TolR also has an N-terminal anchoring domain, a central domain and a periplasmic C-terminal domain which has been proposed to contain an amphipatic helix interacting transiently with the third TM segment of TolQ (Journet et al., 1999; Lazzaroni et al., 1995).

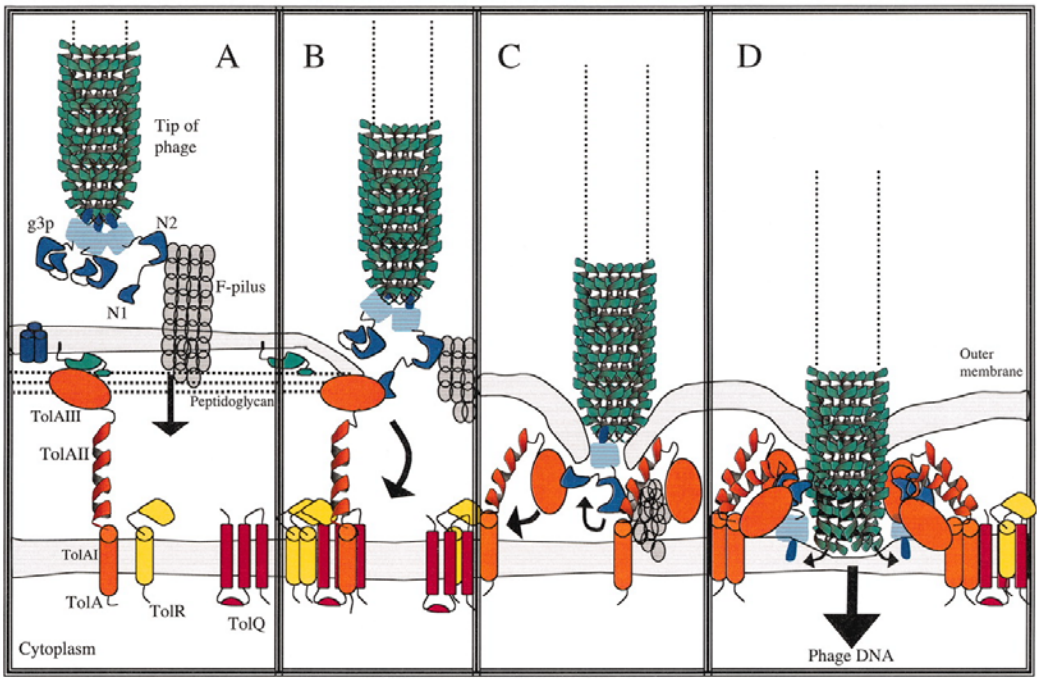


Figure 1-4. Model for the early events in the phage infection of *E. coli*. (A) The N2 domain of pIII interacts with the F-pilus on the outside of the bacteria. The outer membrane protein OmpF (blue cylinders) and Pal lipoprotein (greenish) are also depicted, as they are believed to interact with TolA prior to infection (Cascales et al., 2000; Derouiche et al., 1996; Lazdunski et al., 1998). (B) After pilus retraction, the N1 domain of pIII can bind to the C-terminal domain of TolA (TolAIII). (C) The retracting pilus brings pIII domains in closer contacts with TolA domains. The TolA assumes a more compact state of assembly, which might help bringing the outer and inner membranes of the bacteria close together (Baron et al., 2002). (D) The CT domain of pIII is inserted into the inner membrane, thus creating a pore for phage DNA injection into the bacteria. For further details, see text. Picture taken from Karlsson et al. (2003).

The stoichiometry of each of the Tol-Pal proteins is not well defined. Based on current evidence, a model for the membrane-embedded portion of the complex containing $TolQ_4TolR_2TolA_1$ has been proposed (Cascales et al., 2001). Recently, energy-dependant conformational changes in TolA have been characterized. They depend on the transmembrane domain of TolA and of TolQ and TolR proteins (Germon et al., 2001).
The transmembrane fragment of TolR and the third transmembrane fragment of TolQ are involved in the proton motive force-dependent activation of TolA (Cascales et al., 2000). Genetic suppression studies are consistent with the fact that the Tol system appears to use an ion potential from the inner membrane to induce a conformational change of the C-terminus of TolA in the periplasm, leading to an interaction between inner and outer membrane proteins. This kind of signaling may contribute to the role of the Tol system in maintaining the outer membrane stability. TolQ and TolR present structural and functional homologies not only with ExbB and ExbD of the Ton system but also with MotA and MotB

(see Chapter 1.3), components of the flagellar motor, raising the attractive hypothesis that the transmembrane domains of TolQ, TolR and TolA constitute an ion potential-driven molecular motor (Cascales et al., 2001).

Infection of *E. coli* by a filamentous phage is a multistep process requiring interactions of the phage pIII protein with the F conjugative pilus and the TolQRA system. The F-pilus elaborated by F^+ strains of *E. coli* is an extra-cellular filamentous structure with an outside diameter of 8 nm and a 2 nm central, hydrophilic lumen. It belongs to the so-called “type IV secretion system” (T4SS), which mediates several processes such as bacterial motility, host cell adhesion and natural transformation (Koebnik, 2001). The F-pilus can be very variable in length, ranging from 1 to 20 μm, and is assembled from multiple copies of its structural protein called pilin. Proteins required for its structure, assembly and disassembly are encoded by genes in the “tra” operon on the F conjugative plasmid (Lawley et al., 2003; Manchak et al., 2002). Propilin, encoded by the *traA* gene, is processed to a mature pilin (70 residues) by the *traQ* protein, proposed to act as a specific chaperone (Harris et al., 1999). Originally, the F pilus was shown to be required for the conjugal transfer of the F plasmid DNA from a donor cell into a recipient bacterium lacking the plasmid (Silverman, 1997). This horizontal DNA transfer, called conjugation, is initiated by interaction of the tip of the F pilus with the envelope of the recipient bacterium. By pilus retraction, the donor and recipient cells get closer, thus facilitating the exchange of a single strand of the F plasmid DNA. This process results in the presence of the F plasmid in both the donor and recipient bacteria. It has been pointed out that conjugative F pili morphologically resemble phages made in the absence of pIII, reinforcing the intuitive idea that the two are likely to be evolutionarily related (Rakonjac and Model, 1998).

For reviews about bacterial conjugation see Llosa et al. (2002), about Tra proteins and pilus biogenesis see Kalkum et al. (2002).

Phage infection is initiated by the binding of the N2 domain of pIII to the tip of the F-pilus, which is believed to be present in two to three copies per cell. The exact binding epitope on N2 has been mapped by competitive ELISA (Deng and Perham, 2002). Interaction with the pilus is crucially important, as bacteria lacking F-pili can be infected but only at very low frequency (Krebber et al., 1997). Docking of the phage is thought to be followed by retraction of the pilus, bringing the pIII end of the phage near the surface of the bacteria. It is still a matter of debate whether retraction is a result of normal polymerization-depolymerization cycles of the pilus or whether phage attachment triggers this process. The current model suggests that pilus retraction is triggered by the absence of outgrowth, which is dependant upon ATP synthesis and the proton motive force. Thus, retraction could be the consequence of an interruption in the flow of signalling down the pilus, caused by the attachment of either Ff phage or a recipient cell to the pilus tip (Manchak et al., 2002). The force produced by a single pilus is unexpectedly high, and it constitutes one of the strongest molecular motors reported so far. Biophysical measurements estimated that a single pilus can generate forces exceeding 100 pico-Newton (Maier et al., 2002). The binding of N2 to the tip of the F pilus releases the N1 domain, which becomes free to interact with the C-terminal domain of TolA (TolAIII), also referred as the coreceptor (Riechmann and Holliger, 1997). Surface plasmon resonance (BIACORE) measurements have detected a micromolar

affinity between N1 and TolAIII (Karlsson et al., 2003), but still within pIII, binding of N1 to N2 is entropically favored because the two domains are covalently linked. As mentioned above, domain assembly and dissassembly within pIII is a very slow reaction. The very slow domain re-assembly might ensure that, once N1 and N2 became separated, domain closure is blocked long enough that the pilus can retract to bring N1 into close proximity of TolA and allow the two to productively interact (Martin and Schmid, 2003b). From the structure of the TolAIII-N1 complex (Lubkowski et al., 1999), it has become clear that TolAIII binds N1 at a site overlapping that of N2. Thus, during phage binding to the F pilus, the N1 domain must be displaced from the N2 domain and is thus available for binding of the coreceptor. Nevertheless, although the binding sites for N2 and TolA are overlapping, neither structural homology nor even a similarity of secondary structure between the N1-binding interfaces of N2 and TolA is seen. Recently, the CT domain of pIII was found to interact specifically with the N1-N2 domains, leading to a stable compact conformation of pIII at the tip of the phage, which is expected to unravel during infection (Chatellier et al., 1999). The further steps of the phage infection process are unclear. After the formation of the complex between TolAIII and the N1 domain of pIII, infection may proceed by subsequent interaction of this complex first with the periplasmic C-terminal domain of TolR (Kampfenkel and Braun, 1993; Muller et al., 1993), and then with the entire TolQRA system. Through the action of TolA, the major capsid protein, pVIII, and probably the pVII and pIX minor coat proteins, then depolymerize and become integrated into the host inner membrane as the single-stranded DNA enters the cytoplasm (Click and Webster, 1998). The relief of the capsid internal pressure, which results from the electrostatic repulsion between the polyelectrolyte molecules, is probably the major force driving the DNA out of the capsid and allowing it to emerge from the filamentous particle and pass through the putative pIII pore (Letellier et al., 1999). Some authors inferred that decapsidation and full penetration of the phage DNA are dependant upon and closely linked to its replication (Marco et al., 1974). It has been further hypothesized that pIII may link the phage DNA to a cellular replicative system in or near the inner membrane, and thus participates in initial replication of the DNA (Jazwinski et al., 1973). In vitro studies support the idea that a trimeric form of pIII displays pore-forming properties, thus creating a channel allowing DNA penetration through the inner membrane, and indeed pIII has been shown to interact with DNA (Marco et al., 1974). The estimated radius of the pIII-pore (0.81 nm) is in agreement with such a function, the radius of the phage DNA helix being 0.85 nm (Glaser-Wuttke et al., 1989). The rate of DNA transport, although varying from one phage to another, can reach value as high as 3’000 base pairs/second, a value significantly higher than that attained during the transport of DNA in conjugation and natural transformation (100 bases/second). Once in the *E. coli* membrane, the pVIII protein of the infecting phage joins the pool of newly synthesized pVIII, and both are assembled to newly formed particles. Electron microscope studies suggest that a chloroform/water interface is able to induce morphological changes, contracting the filaments sequentially into shortened forms (I-forms), and then into spheroidal particles (S-forms). Chloroform might lead to phage contraction, due to its interactions with pVIII. The conversion to the I-form could correspond to the activation step that enables depolymerization of pVIII into the bilayer, while the I-form to S-form step might mimic the depolymerization step of the infection process (Oh et al., 1999).

1.3 TonB-ExbB-ExbD system and its analogy to TolQRA

Colicins are bacterial protein toxins, encoded by plasmids, which are active against *E. coli* and other related species. They are produced in large amounts and are generally released into the extracellular medium. A few colicins have been structurally characterized (Cheng et al., 2002; Elkins et al., 1997; Hilsenbeck et al., 2004; Soelaiman et al., 2001; Sui et al., 2002; Vetter et al., 1998; Wiener et al., 1997). They use different ways to kill susceptible cells, ranging from the ability to depolarize the inner membrane or to cytotoxic activity against cytoplasmic nucleic acids (Lazdunski et al., 2000; Lazdunski et al., 1998). The mechanism by which these folded proteins are able to cross the membrane of the target cells has been extensively studied. To enter bacterial cells, colicins parasitize outer membrane receptors used by sensitive cells for important biological functions, such as porins (e.g. OmpF), vitamin B12 receptor (BtuB) or high-affinity receptors for iron (FepA, FhuA, FecA), then they interact with proteins in the periplasmic space to reach their target. Each colicin shows the same type of organization, coherent with their mechanism of action. Their structure comprises three distinct domains; (i) a central domain involved in the recognition of a specific receptor at the surface of the target cell, (ii) an N-terminal domain involved in translocation, and (iii) a C-terminal domain responsible for lethal activity. Both TonB and Tol-Pal systems have been parasitized and are required for the import through the cell envelope of both colicins and phage DNA.

Interestingly, in a way similar to colicins, a specific interaction or action can be assigned to each of the domains of pIII. The central domain (N2) interacts with the F pilus playing the role of receptor. Together with a retraction of the pilus, this unmasks the N1 domain that can then interact in the periplasm with the C-terminal domain of TolA (TolAIII), which plays the role of a co-receptor. The N-terminal domain of pIII is comparable to the translocation domain of colicins, and indeed, it can play the role of a colicin translocation domain since it was shown that a fusion of the N-terminal domains of pIII (N1 + N2) to the catalytic domain of colicin E3 was active against bacteria (Jakes et al., 1988). Colicins are classified into two groups depending on the translocation system they use. Group A colicins need the Tol proteins, group B colicins use the TonB-ExbB-ExbD system.

For reviews about colicins, see Cao and Klebba (2002), Lazdunski et al. (1998), Zakharov and Cramer (2002).

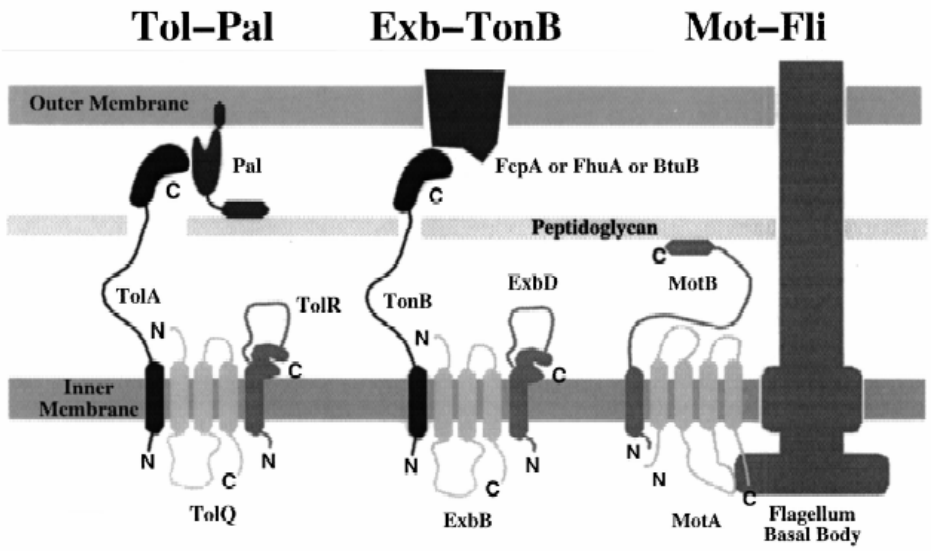


Figure 1-5. Schematic illustration of the different components of the three ion-potential coupled systems discussed in the text: the Tol-Pal system (left), the TonB/ExbB/ExbD system (centre) and the flagellar motor MotAB (right). Picture taken from Cascales et al. (2001).

Some components of the Tol system and the TonB system are homologous (Figure 1-5). Homologues of TonB have been found widely distributed among Gram-negative bacteria. They energize the active uptake of iron obtained from mammalian iron-binding proteins (Braun and Braun, 2002), iron siderophores (Braun and Killmann, 1999) and of vitamin B12 (Buchanan et al., 1999; Cadieux and Kadner, 1999) across the outer membrane. TonB-ExbB-ExbD are three components of the multimeric TonB-dependent energy-transducing system (Higgs et al., 1998; Larsen et al., 1999) present in stoichiometries of about 1:7:2 (Higgs et al., 2002). According to the available experimental evidence, the energized TonB protein physically contacts a ligand-bound outer membrane receptor and transduces energy from ExbB-ExbD, inducing a conformational change in the outer membrane receptor (Ferguson et al., 1998; Lazzaroni et al., 1999; Letain and Postle, 1997). This results in the release of the ligand into the periplasm. The two Exb proteins together contribute to the stability of TonB, therefore these three proteins are presumed to interact directly through their transmembrane domains. An essential function has been ascribed to ExbD, and in particular to aspartate 25 and leucine 132 within the protein. It is conceivable that aspartate 25 of ExbD and histidine 20 of TonB are involved in sensing the electrochemical potential of the inner membrane, which induces a conformation in the Ton complex that triggers the opening of the channels in the outer membrane receptor proteins (Braun, 1995).

TolQ and TolR are structurally and functionally related to ExbB and ExbD, respectively (Braun and Herrmann, 1993). TolA and TonB have a similar elongated conformation but are homologous only in their N-terminal transmembrane domains, where they both have a well conserved “SHLS” motif (Germon et al., 1998; Karlsson et al., 1993).

The TolQ-TolR-TolA-N-terminus complex can partially replace ExbB-ExbD-TonB-N-terminus for energy transduction (Karlsson et al., 1993). However, the TolA and TonB proteins span the periplasm by quite different mechanisms, since the TolA central domain has an α–helical structure whereas that of TonB has a stretch of X-Pro repeats. Thus the presence of a similar trans-periplasmic system encoded by the Tol genes (TolQ and TolR) that can compensate for mutations in two of the inner membrane proteins of the TonB system (ExbB and ExbD) suggests that a similar mode of energy transduction from the inner membrane to the outer membrane is used by both the Ton and Tol systems. TolQ-TolR seem to transduce the proton-motive force to TolA by a mechanism probably homologous to the transduction of the proton-motive force to TonB by ExbB-ExbD. The TolA protein might be involved in transducing energy from the inner membrane to the Pal lipoprotein or via Pal lipoprotein to some other periplasmic or outer membrane component. TolA could drive newly synthesized outer membrane components across the periplasm through its C-terminal interaction with Pal (Levengood et al., 1991). It has been suggested that the Tol proteins are involved in the dynamic assembly of outer membrane trimeric porins (Derouiche et al., 1996). Thus, the TonB and Tol-Pal systems are both able to couple the cytoplasmic membrane proton gradient to energy-requiring processes (Cascales et al., 2000) and thus energize active transport across the outer membrane and maintain cell envelope integrity, respectively. The proton motive force corresponds to an electrochemical potential of protons, resulting from the active transport of protons across the inner membrane into the cytoplasm of *E. coli*. It is assumed that the two systems probably originated from a common ancestor, and their physiological functions appear to be essential because they are conserved in many Gram-negative bacteria.

The importance of the proton motive force in the phage infection mechanism still remains largely elusive: depolarization of the inner membrane seems to be essential for successful infection of a bacterial cell with phage DNA. In several investigations it has been demonstrated that phage infection results in a reduction of the proton-motive force and in an exchange of cations across the cell membrane (Feucht et al., 1990).

At first sight, these Ton and Tol systems have little in common with the flagellar motor other than the use of a transmembrane electrochemical gradient as an energy source. The bacterial flagellum is a complex macromolecular assemblage forming a multipartite structure composed of a long helical propeller, a flexible hook region and a rotary motor in the bacterial inner membrane. The source of energy for motor rotation is the membrane gradient of protons or, in some species, sodium ions (Okabe et al., 2002). The flagellar motor thus converts chemical energy into mechanical work. MotA and MotB are integral membrane proteins essential for flagellar rotation, which are believed to form the stator of the motor (Figure 1-5). They bind to each other and function together to conduct protons across the inner membrane. Two conserved proline residues (Pro 173 and Pro 222) of MotA are important for torque generation. Asp 32 of MotB is also critical for function, because it forms part of the proton path through the motor. All three residues are predicted to be near the cytoplasmic ends of membrane segments, so they could very likely function directly in coupling proton transfer to motor rotation (Braun et al., 1999). There are presently no high-resolution structural data on the MotA/MotB complexes, but a new model of the flagellar motor rotation has been proposed, based on KcsA potassium ion channel dynamics

(Schmitt, 2003). Biochemical data determined a stoichiometry of four copies of MotA for two copies of MotB (Blair, 2003). Based on sequence alignment and notably the conservation of a critical aspartate residue in the TM helix of ExbB, TolR and MotB, it has been postulated that the TolQRA complex could exist in the inner membrane with the same stoichiometry. Therefore, a model was proposed for the membrane-embedded portion of the complex, composed of $TolQ_4TolR_2TolA_1$. This has raised the attractive idea that the last two transmembrane (TM) helices of TolQ, the single TM helix of TolR and possibly its C-terminus might form an ion pore (Cascales et al., 2001). Molecular dynamics simulations were then used to propose a simple structure of MotA-MotB. This theoretical structure served as a template to design a structural model for the evolutionary-related ExbB-ExbD-TonB system. It allows to identify conserved residues in the complex (Figure 1-6); it is remarkable that all of them but one (F34) face inward and may be important for function. From this model, a pathway for proton transport in ExbB-ExbD-TonB implicating water molecules is suggested. A possible role of TonB in this pathway is suggested and discussed (Zhai et al., 2003).

For reviews about the bacterial flagellar motor, see DeRosier (1998) and Blair (2003).

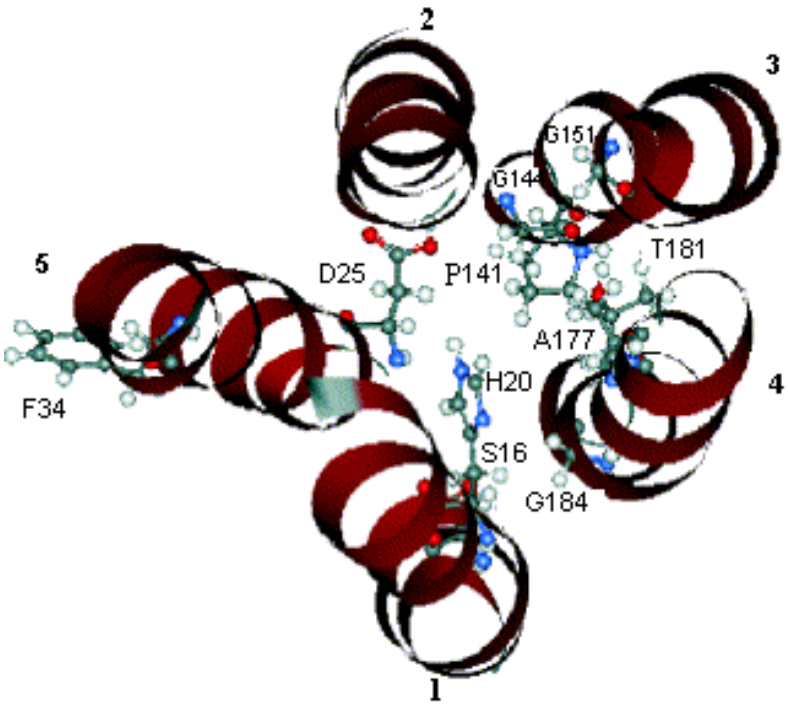


Figure 1-6. This figure shows the conserved transmembrane segments of the ExbB-ExbD-TonB complex. These residues are P141, G144 and G151 on the 2nd helix of ExbB; A177, T181 and G184 on the 3rd helix of ExbB, D25 and F34 on the helix of ExbD, and H20 and S16 associated with the helix of TonB. T181 and D25 might be involved in proton translocation. Picture taken from Zhai et al. (2003).

1.4 Display of proteins on filamentous bacteriophages

The major interactions that occur among the coat proteins (pIII and pVIII) during the assembly process involve the hydrophobic regions of these proteins, which initially span the inner membrane. These membrane-associated assembly properties of the coat proteins allow packaging of chimeric proteins into the phage particle. Any protein fused to the periplasmic portion of these coat proteins should theoretically have a good chance of being packaged into a phage particle, provided it can be translocated efficiently across the inner membrane and not interfere with the process that occurs at the assembly site. In 1985, George Smith demonstrated that a foreign peptide could be displayed on the surface of filamentous bacteriophage as a genetic fusion with the minor coat protein, pIII (Scott and Smith, 1990; Smith, 1985). Because the displayed peptide forms a structural association with its encoding DNA, phage display was immediately recognized as a powerful tool for selecting and evolving ligands from large combinatorial librairies. In this way, a physical linkage between phenotype (displayed peptide) and genotype (DNA) is established, and extremely diverse libraries (>10^{11} members) of DNA-encoded peptides or proteins can be generated with simple molecular biology methods. Furthermore, phage-displayed libraries can be amplified by passage through a bacterial host, and even the largest repertoires can be contained in less than one millilitre of solution. The bacteria tolerate this process quite well and continue to grow and divide with a generation time only about 50% longer than that of uninfected bacteria. There is a burst of about 1000 phage particles produced in the first generation after infection, and then the bacteria produce about 100-200 phage particles per generation. This continues for many generations, resulting in titers of 10^{11} to 10^{12} particles per ml. The plaques are turbid and of varying size and contain about 10^{8} infective phage particles.

By using selection with immobilized ligands (on plates or magnetic beads), library pools can be enriched for polypeptides with particular binding characteristics (Figure 1-7). Following binding selection, individual clones from enriched pools can be analyzed by using enzyme-linked immunosorbant assays (ELISA) to quantify binding in a high-throughput fashion. Finally, and most importantly, the amino acid sequence of any clone can be readily deciphered by sequencing the DNA inside the phage particle. Historically, phage display has relied on creating amino-terminal fusion peptides with either pIII or pVIII. Small peptides can be displayed on all copies of pVIII in the coat, in contrast to larger proteins, which require a mosaic display with the inclusion of some wildtype pVIII so as not to disrupt assembly. The great majority of phage display libraries fuse foreign peptides and proteins to the N-terminus of either pIII or pVIII; however, recent studies indicate that pVI (Jespers et al., 1995), as well as pVII and pIX may be used for this purpose (Gao et al., 2002; Gao et al., 1999) suggesting greater flexibility in phage particles than previously thought. Using a special two-gene pVIII display system, Sidhu and colleagues showed that the pVIII can actually accommodate a surprisingly large number of mutations. Furthermore, many of these mutations improve packaging, and hence protein display, in a predominantly wild-type coat (Sidhu et al., 2000). As an extreme example of sequence divergence, Weiss and colleagues were able to construct a functional pVIII variant with only 50% identity with the wild-type

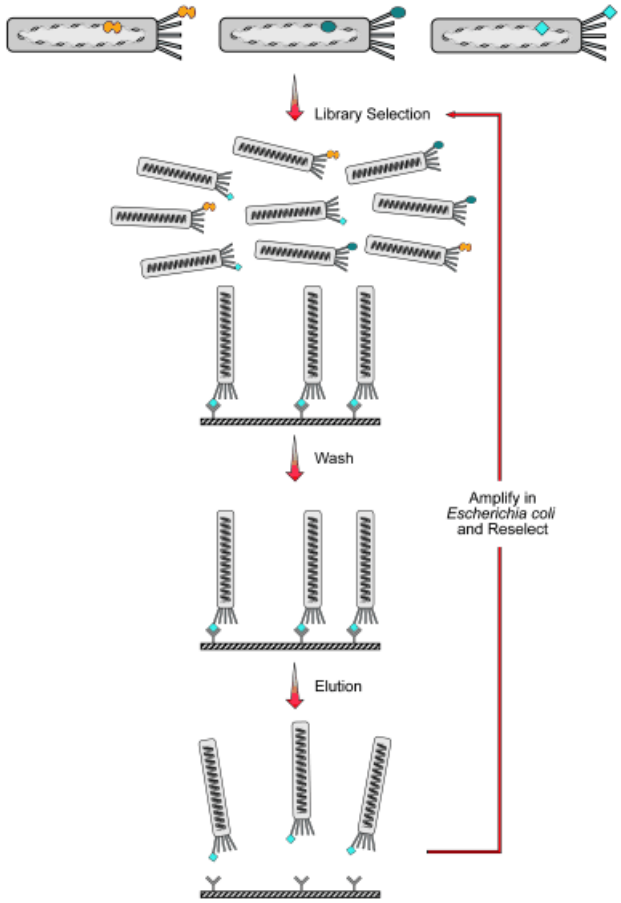


Figure 1-7. In vitro selection with phage display. Proteins are displayed on phage particles that also contain the encoding DNA. Display is usually achieved by fusion to pIII located at one end of the particle, as depicted here. The library is incubated with an immobilized target to select for binders, and non-binding phages are removed by washing. Bound phages are then eluted and amplified in *E. coli*. Amplified phage pools can then be subjected to additional rounds of selection, or alternatively, the sequences of individual clones can be determined by sequencing the encapsulated DNA. Picture taken from Sidhu et al. (2003).

(Weiss et al., 2000). The same group could even evolve pVIII to render it compatible with C-terminal display. Improved variants of pVIII could be identified, which supported the simultaneous display of N and C-terminal fusions (Held and Sidhu, 2004).

Not all proteins are efficiently displayed on phage, yet a wide range of peptides and proteins can be used. In addition, various forms of antibodies, including Fabs and single-chain antibodies (scFvs) have been displayed. Originally, the display of large polypeptides and proteins was severely limited. This problem was circumvented by the use of two-gene phagemid systems which enabled the packaging of fusion gene products in a phage coat predominantly composed of wild-type proteins supplied *in trans* from a helper phage. In such systems, the recombinant pIII fusion-protein is encoded on a phagemid. This phagemid produces large amounts of the recombinant display protein, but is unable to make complete

large phage, unless the bacteria carrying the phagemid become infected by a helper phage which supplies all the other proteins required to assemble functional phage. Thus, the pIII fusion (supplied from the phagemid) is a non-functional coat protein that incorporates into the phage particle as a non-essential addition to the wild-type pIII (supplied from the helper phage). However, as the incorporation of the pIII fusion is slightly disfavoured in comparison to the wild-type pIII, it is assumed that recombinant phage generated with a phagemid system carry only one copy of the pIII fusion along with two to four copies of the wild-type pIII. We speak of a monovalent display of the pIII fusion-protein at the tip of the phage.

pIII is the minor coat protein most widely used for display of proteins (and especially antibodies). Its main disadvantage of displaying only about 5 molecules per phage particle is compensated by the fact that large proteins appear to package reasonably well. In our set-up discussed in the third section, the protein (Fab antibody fragment) is fused to the C-terminal domain of pIII. Once in the membrane of the host bacteria, the foreign protein presumably does not interfere with assembly, a process mediated by the C-terminus of pIII. The ability to display an active protein depends on whether the fusion protein can be translocated properly into the membrane, fold properly, escape degradation in the periplasm, and be acceptable for packaging into phage.

It is rather amazing how versatile and flexible the phage-display system is, considering the possible constraints that the processes of protein translocation, folding, and phage assembly might place on it. Recent developments of this powerful technology have made it an attractive tool for selecting the protein of choice. This great success has led to a variety of clever methods designed in order to extend the range of future applications. The potential of phages as an *in vivo* or *in vitro* targeted gene delivery vehicle was illustrated by the re-engineering of phage F5, in order to infect eukaryotic cells (Poul and Marks, 1999). It was recently demonstrated that a modified phage-display vector can be used to specifically target mammalian cells, thus being able to deliver genes into eukaryotic cells. The principles of directed evolution used in phage display could be applied to phage vectors to tailor them for gene delivery to specific target cells *in vivo* (Larocca and Baird, 2001; Larocca et al., 2001).

1.5 Work outline

There are only few ways, either natural or engineered, that new binding entities can be discovered in a short period of time. The immune system is probably the best known natural example. When challenged with a foreign antigen, high affinity antibodies are generated in a matter of weeks. This process occurs in stages, with the first step being the generation of low-affinity antibodies, displayed on the surface of B-cells, with each cell presenting a single antibody species. Initially, there is a resident library of naïve antibodies (believed to be in the range of 10^7 different molecules) that cover a wide range of diversity. At this point, the affinities are only in the micromolar range; but the binding properties of these antibodies of the primary response are boosted significantly by multiple attachment to the antigen (often in multimeric form). This avidity effect, coupled to the large number of naïve antibodies in the starting pool, allow a small group of antibodies to be clonally selected. The second stage in antibody selection involves an affinity maturation process. Mutations are introduced into the variable domains of selected antibodies by a process known as “somatic hypermutation”. It results in the production of antibodies with more diversity, and permits improvements in the complementarity between the antibody and antigen. The affinities of these natural antibodies are increased; they are also called “antibodies of the secondary response”.

Phage display, as it is practised today for selection of binders from naïve libraries, mirrors the natural immune system (Marks et al., 1992). One begins with a large and diverse set of proteins (in our case, antibody fragments) that are presented in a polyvalent format at the tip of the phage particle. Antibody fragments, selected from this polyvalent format bind their target in the high micromolar range. In a subsequent step, antibody fragments with higher affinities are generated from these leads by introducing additional random mutations and transferring them onto a lower valency format to allow for selection of antibodies with affinities in the nanomolar range.

The present work aims at understanding the structural processes governing bacterial infection by a filamentous bacteriophage, which is one of the foundations of phage display. The next section of this work (section 2) is devoted to the study of three proteins that are intricately implicated in the infection process. First, an extensive study of the C-terminus of pIII provided some data, hinting at its intrinsic unstability in solution; we were unable to get this fragment of pIII in its full form. We provide evidence that once isolated from inclusion bodies in *E. coli*, it gets proteolytically processed, in a spontaneous way and by some still unknown proteases. The primary goal of obtaining a crystal structure of this domain thus failed. TolQRA and one member of this complex is the topic of the next chapter. The inner membrane-anchored TolR has been shown to be critical for phage entry into *E. coli*. The biochemical and spectroscopic characterization of its periplasmic C-terminal domain showed it to possess a high degree of conformational flexibility. This observation was corroborated by the study of several constructs of various lengths, which all show a non-optimal dispersion in two-dimensional NMR spectra, thus preventing any further structural characterization. Finally, we will draw our attention to the related ExbB-ExbD-TonB system. TonB is indirectly used by colicins and siderophores to get entry into bacterial cells, but this pathway has been

shown to be parasitized by certain phages as well. The transport mechanism of siderophores into *E. coli* is discussed in the light of our published crystal structure of the C-terminal domain of TonB. Our findings are confronted with new recent results on TonB itself, as well as on crystal structures of outer membrane transporters which are contacted by the C-terminus of TonB.

The third and last section illustrates the use of phage display to select an antibody fragment with better binding properties toward an antigen of immunological relevance. Based on earlier work in our laboratory (Frédéric Pecorari, Maurice Brozzo, Stefan Ewert, all unpublished results), we tried to improve the dissociation constant of a Fab antibody fragment directed against the MHC-gp33 peptide complex. We have applied phage display to affinity mature this Fab fragment and isolated novel Fab clones against the MHC-gp33 from a self-constructed antibody library. Interestingly, the paratope of the improved clones could be mapped exclusively to the heavy-chain of the original Fab. A single domain antibody was constructed consisting solely of the variable domain of this heavy chain. We provide evidence that it retains full binding activity towards the MHC-gp33, despite the fact that the binding affinity did not show substantial improvement. This finding has fundamental implications with regard to the affinity maturation process and the structural basis of the peptide-MHC recognition mechanism.

1.6 Literature

Armstrong, J., Perham, R. N., and Walker, J. E. (1981). Domain structure of bacteriophage fd adsorption protein. FEBS Lett 135, 167-172.

Baron, C., D, O. C., and Lanka, E. (2002). Bacterial secrets of secretion: EuroConference on the biology of type IV secretion processes. Mol Microbiol 43, 1359-1365.

Bernadac, A., Gavioli, M., Lazzaroni, J. C., Raina, S., and Lloubes, R. (1998). Escherichia coli tol-pal mutants form outer membrane vesicles. J Bacteriol 180, 4872-4878.

Blair, D. F. (2003). Flagellar movement driven by proton translocation. FEBS Lett 545, 86-95.

Braun, T. F., Poulson, S., Gully, J. B., Empey, J. C., Van Way, S., Putnam, A., and Blair, D. F. (1999). Function of proline residues of MotA in torque generation by the flagellar motor of Escherichia coli. J Bacteriol 181, 3542-3551.

Braun, V. (1995). Energy-coupled transport and signal transduction through the gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. FEMS Microbiol Rev 16, 295-307.

Braun, V., and Braun, M. (2002). Active transport of iron and siderophore antibiotics. Curr Opin Microbiol 5, 194-201.

Braun, V., and Herrmann, C. (1993). Evolutionary relationship of uptake systems for biopolymers in Escherichia coli: cross-complementation between the TonB-ExbB-ExbD and the TolA-TolQ-TolR proteins. Mol Microbiol 8, 261-268.

Braun, V., and Killmann, H. (1999). Bacterial solutions to the iron-supply problem. Trends Biochem Sci 24, 104-109.

Buchanan, S. K., Smith, B. S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D., and Deisenhofer, J. (1999). Crystal structure of the outer membrane active transporter FepA from Escherichia coli. Nat Struct Biol 6, 56-63.

Cadieux, N., and Kadner, R. J. (1999). Site-directed disulfide bonding reveals an interaction site between energy-coupling protein TonB and BtuB, the outer membrane cobalamin transporter. Proc Natl Acad Sci USA 96, 10673-10678.

Cao, Z., and Klebba, P. E. (2002). Mechanisms of colicin binding and transport through outer membrane porins. Biochimie 84, 399-412.

Cascales, E., Gavioli, M., Sturgis, J. N., and Lloubes, R. (2000). Proton motive force drives the interaction of the inner membrane TolA and outer membrane pal proteins in Escherichia coli. Mol Microbiol 38, 904-915.

Cascales, E., Lloubes, R., and Sturgis, J. N. (2001). The TolQ-TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA-MotB. Mol Microbiol 42, 795-807.

Chatellier, J., Hartley, O., Griffiths, A. D., Fersht, A. R., Winter, G., and Riechmann, L. (1999). Interdomain interactions within the gene 3 protein of filamentous phage. FEBS Lett 463, 371-374.

Cheng, Y. S., Hsia, K. C., Doudeva, L. G., Chak, K. F., and Yuan, H. S. (2002). The crystal structure of the nuclease domain of colicin E7 suggests a mechanism for binding to double-stranded DNA by the H-N-H endonucleases. J Mol Biol 324, 227-236.

Click, E. M., and Webster, R. E. (1998). The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. J Bacteriol 180, 1723-1728.

Crissman, J. W., and Smith, G. P. (1984). Gene-III protein of filamentous phages: evidence for a carboxyl- terminal domain with a role in morphogenesis. Virology 132, 445-455.

Deng, L. W., Malik, P., and Perham, R. N. (1999). Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. Virology 253, 271-277.

Deng, L. W., and Perham, R. N. (2002). Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. J Mol Biol 319, 603-614.

DeRosier, D. J. (1998). The turn of the screw: the bacterial flagellar motor. Cell 93, 17-20.

Derouiche, R., Benedetti, H., Lazzaroni, J. C., Lazdunski, C., and Lloubes, R. (1995). Protein complex within Escherichia coli inner membrane. TolA N-terminal domain interacts with TolQ and TolR proteins. J Biol Chem 270, 11078-11084.

Derouiche, R., Gavioli, M., Benedetti, H., Prilipov, A., Lazdunski, C., and Lloubes, R. (1996). TolA central domain interacts with Escherichia coli porins. Embo J 15, 6408-6415.

Elkins, P., Bunker, A., Cramer, W. A., and Stauffacher, C. V. (1997). A mechanism for toxin insertion into membranes is suggested by the crystal structure of the channel-forming domain of colicin E1. Structure 5, 443-458.

Endemann, H., and Model, P. (1995). Location of filamentous phage minor coat proteins in phage and in infected cells. J Mol Biol 250, 496-506.

Feng, J. N., Russel, M., and Model, P. (1997). A permeabilized cell system that assembles filamentous bacteriophage. Proc Natl Acad Sci USA 94, 4068-4073.

Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (1998). Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science 282, 2215-2220.

Feucht, A., Schmid, A., Benz, R., Schwarz, H., and Heller, K. J. (1990). Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. J Biol Chem 265, 18561-18567.

Gailus, V., Ramsperger, U., Johner, C., Kramer, H., and Rasched, I. (1994). The role of the adsorption complex in the termination of filamentous phage assembly. Res Microbiol 145, 699-709.

Gao, C., Mao, S., Ditzel, H. J., Farnaes, L., Wirsching, P., Lerner, R. A., and Janda, K. D. (2002). A cell-penetrating peptide from a novel pVII-pIX phage-displayed random peptide library. Bioorg Med Chem 10, 4057-4065.

Gao, C., Mao, S., Lo, C. H., Wirsching, P., Lerner, R. A., and Janda, K. D. (1999). Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. Proc Natl Acad Sci USA 96, 6025-6030.

Germon, P., Clavel, T., Vianney, A., Portalier, R., and Lazzaroni, J. C. (1998). Mutational analysis of the Escherichia coli K-12 TolA N-terminal region and characterization of its TolQ-interacting domain by genetic suppression. J Bacteriol 180, 6433-6439.

Germon, P., Ray, M. C., Vianney, A., and Lazzaroni, J. C. (2001). Energy-dependent conformational change in the TolA protein of Escherichia coli involves its N-terminal domain, TolQ, and TolR. J Bacteriol 183, 4110-4114.

Glaser-Wuttke, G., Keppner, J., and Rasched, I. (1989). Pore-forming properties of the adsorption protein of filamentous phage fd. Biochim Biophys Acta 985, 239-247.

Gray, C. W., Brown, R. S., and Marvin, D. A. (1981). Adsorption complex of filamentous fd virus. J Mol Biol 146, 621-627.

Guihard, G., Boulanger, P., Benedetti, H., Lloubes, R., Besnard, M., and Letellier, L. (1994). Colicin A and the Tol proteins involved in its translocation are preferentially located in the contact sites between the inner and outer membranes of Escherichia coli cells. J Biol Chem 269, 5874-5880.

Harris, R. L., Sholl, K. A., Conrad, M. N., Dresser, M. E., and Silverman, P. M. (1999). Interaction between the F plasmid TraA (F-pilin) and TraQ proteins. Mol Microbiol 34, 780-791.

Held, H. A., and Sidhu, S. S. (2004). Comprehensive mutational analysis of the M13 major coat protein: improved scaffolds for C-terminal phage display. J Mol Biol 340, 587-597.

Higgs, P. I., Larsen, R. A., and Postle, K. (2002). Quantification of known components of the Escherichia coli TonB energy transduction system: TonB, ExbB, ExbD and FepA. Mol Microbiol 44, 271-281.

Higgs, P. I., Myers, P. S., and Postle, K. (1998). Interactions in the TonB-dependent energy transduction complex: ExbB and ExbD form homomultimers. J Bacteriol 180, 6031-6038.

Hilsenbeck, J. L., Park, H., Chen, G., Youn, B., Postle, K., and Kang, C. (2004). Crystal structure of the cytotoxic bacterial protein colicin B at 2.5 A resolution. Mol Microbiol 51, 711-720.

Holliger, P., and Riechmann, L. (1996). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5, 265-275.

Holliger, P., Riechmann, L., and Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 A: evidence for conformational lability. J Mol Biol 288, 649-657.

Houbiers, M. C., Spruijt, R. B., Wolfs, C. J., and Hemminga, M. A. (1999). Conformational and aggregational properties of the gene 9 minor coat protein of bacteriophage M13 in membrane-mimicking systems. Biochemistry 38, 1128-1135.

Jakes, K. S., Davis, N. G., and Zinder, N. D. (1988). A hybrid toxin from bacteriophage f1 attachment protein and colicin E3 has altered cell receptor specificity. J Bacteriol 170, 4231-4238.

Jazwinski, S. M., Marco, R., and Kornberg, A. (1973). A coat protein of the bacteriophage M13 virion participates in membrane- oriented synthesis of DNA. Proc Natl Acad Sci USA 70, 205-209.

Jespers, L. S., Messens, J. H., De Keyser, A., Eeckhout, D., Van den Brande, I., Gansemans, Y. G., Lauwereys, M. J., Vlasuk, G. P., and Stanssens, P. E. (1995). Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. Biotechnology (N Y) 13, 378-382.

Journet, L., Rigal, A., Lazdunski, C., and Benedetti, H. (1999). Role of TolR N-terminal, central, and C-terminal domains in dimerization and interaction with TolA and tolQ. J Bacteriol 181, 4476-4484.

Kalkum, M., Eisenbrandt, R., Lurz, R., and Lanka, E. (2002). Tying rings for sex. Trends Microbiol 10, 382.

Kampfenkel, K., and Braun, V. (1993). Membrane topologies of the TolQ and TolR proteins of Escherichia coli: inactivation of TolQ by a missense mutation in the proposed first transmembrane segment. J Bacteriol 175, 4485-4491.

Karlsson, F., Borrebaeck, C. A., Nilsson, N., and Malmborg-Hager, A. C. (2003). The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains. J Bacteriol 185, 2628-2634.

Karlsson, M., Hannavy, K., and Higgins, C. F. (1993). A sequence-specific function for the N-terminal signal-like sequence of the TonB protein. Mol Microbiol 8, 379-388.

Koebnik, R. (2001). The role of bacterial pili in protein and DNA translocation. Trends Microbiol 9, 586-590.

Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L., and Plückthun, A. (1997). Selectively-infective phage (SIP): a mechanistic dissection of a novel in vivo selection for protein-ligand interactions. J Mol Biol 268, 607-618.

Larocca, D., and Baird, A. (2001). Receptor-mediated gene transfer by phage-display vectors: applications in functional genomics and gene therapy. Drug Discov Today 6, 793-801.

Larocca, D., Jensen-Pergakes, K., Burg, M. A., and Baird, A. (2001). Receptor-targeted gene delivery using multivalent phagemid particles. Mol Ther 3, 476-484.

Larsen, R. A., Thomas, M. G., and Postle, K. (1999). Protonmotive force, ExbB and ligand-bound FepA drive conformational changes in TonB. Mol Microbiol 31, 1809-1824.

Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. FEMS Microbiol Lett 224, 1-15.

Lazdunski, C., Bouveret, E., Rigal, A., Journet, L., Lloubes, R., and Benedetti, H. (2000). Colicin import into Escherichia coli cells requires the proximity of the inner and outer membranes and other factors. Int J Med Microbiol 290, 337-344.

Lazdunski, C. J., Bouveret, E., Rigal, A., Journet, L., Lloubes, R., and Benedetti, H. (1998). Colicin import into Escherichia coli cells. J Bacteriol 180, 4993-5002.

Lazzaroni, J. C., Germon, P., Ray, M. C., and Vianney, A. (1999). The Tol proteins of Escherichia coli and their involvement in the uptake of biomolecules and outer membrane stability. FEMS Microbiol Lett 177, 191-197.

Lazzaroni, J. C., Vianney, A., Popot, J. L., Benedetti, H., Samatey, F., Lazdunski, C., Portalier, R., and Geli, V. (1995). Transmembrane alpha-helix interactions are required for the functional assembly of the Escherichia coli Tol complex. J Mol Biol 246, 1-7.

Letain, T. E., and Postle, K. (1997). TonB protein appears to transduce energy by shuttling between the cytoplasmic membrane and the outer membrane in Escherichia coli. Mol Microbiol 24, 271-283.

Letellier, L., Plançon, L., Bonhivers, M., and Boulanger, P. (1999). Phage DNA transport across membranes. Res Microbiol 150, 499-505.

Levengood, S. K., Beyer, W. F., Jr., and Webster, R. E. (1991). TolA: a membrane protein involved in colicin uptake contains an extended helical region. Proc Natl Acad Sci USA 88, 5939-5943.

Llosa, M., Gomis-Ruth, F. X., Coll, M., and Cruz Fd Fde, L. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol 45, 1-8.

Lopez, J., and Webster, R. E. (1985). Assembly site of bacteriophage f1 corresponds to adhesion zones between the inner and outer membranes of the host cell. J Bacteriol 163, 1270-1274.

Lubkowski, J., Hennecke, F., Plückthun, A., and Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nat Struct Biol 5, 140-147.

Lubkowski, J., Hennecke, F., Plückthun, A., and Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure Fold Des 7, 711-722.

Maier, B., Potter, L., So, M., Seifert, H. S., and Sheetz, M. P. (2002). Single pilus motor forces exceed 100 pN. Proc Natl Acad Sci USA 99, 16012-16017.

Makowski, L. (1992). Terminating a macromolecular helix. Structural model for the minor proteins of bacteriophage M13. J Mol Biol 228, 885-892.

Manchak, J., Anthony, K. G., and Frost, L. S. (2002). Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer. Mol Microbiol 43, 195-205.

Marassi, F. M., and Opella, S. J. (2003). Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints. Protein Sci 12, 403-411.

Marco, R., Jazwinski, S. M., and Kornberg, A. (1974). Binding, eclipse, and penetration of the filamentous bacteriophage M13 in intact and disrupted cells. Virology 62, 209-223.

Marks, J. D., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1992). Molecular evolution of proteins on filamentous phage. Mimicking the strategy of the immune system. J Biol Chem 267, 16007-16010.

Martin, A., and Schmid, F. X. (2003a). Evolutionary stabilization of the gene-3-protein of phage fd reveals the principles that govern the thermodynamic stability of two-domain proteins. J Mol Biol 328, 863-875.

Martin, A., and Schmid, F. X. (2003b). The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3 protein of phage fd. J Mol Biol 329, 599-610.

Martin, A., and Schmid, F. X. (2003c). A proline switch controls folding and domain interactions in the gene-3-protein of the filamentous phage fd. J Mol Biol 331, 1131-1140.

Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. Curr Opin Struct Biol 8, 150-158.

Marvin, D. A., Hale, R. D., Nave, C., and Helmer-Citterich, M. (1994). Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, f1, M13), If1 and IKe. J Mol Biol 235, 260-286.

Muller, M. M., Vianney, A., Lazzaroni, J. C., Webster, R. E., and Portalier, R. (1993). Membrane topology of the Escherichia coli TolR protein required for cell envelope integrity. J Bacteriol 175, 6059-6061.

Oh, J. S., Davies, D. R., Lawson, J. D., Arnold, G. E., and Dunker, A. K. (1999). Isolation of chloroform-resistant mutants of filamentous phage: localization in models of phage structure. J Mol Biol 287, 449-457.

Okabe, M., Yakushi, T., Kojima, M., and Homma, M. (2002). MotX and MotY, specific components of the sodium-driven flagellar motor, colocalize to the outer membrane in Vibrio alginolyticus. Mol Microbiol 46, 125-134.

Poul, M. A., and Marks, J. D. (1999). Targeted gene delivery to mammalian cells by filamentous bacteriophage. J Mol Biol 288, 203-211.

Rakonjac, J., Feng, J., and Model, P. (1999). Filamentous phage are released from the bacterial membrane by a two- step mechanism involving a short C-terminal fragment of pIII. J Mol Biol 289, 1253-1265.

Rakonjac, J., and Model, P. (1998). Roles of pIII in filamentous phage assembly. J Mol Biol 282, 25-41.

Riechmann, L., and Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. Cell 90, 351-360.

Russel, M., Linderoth, N. A., and Sali, A. (1997). Filamentous phage assembly: variation on a protein export theme. Gene 192, 23-32.

Schmitt, R. (2003). Helix rotation model of the flagellar rotary motor. Biophys J 85, 843-852.

Scott, J. K., and Smith, G. P. (1990). Searching for peptide ligands with an epitope library. Science 249, 386-390.

Sidhu, S. S., Fairbrother, W. J., and Deshayes, K. (2003). Exploring protein-protein interactions with phage display. Chembiochem 4, 14-25.

Sidhu, S. S., Weiss, G. A., and Wells, J. A. (2000). High copy display of large proteins on phage for functional selections. J Mol Biol 296, 487-495.

Silverman, P. M. (1997). Towards a structural biology of bacterial conjugation. Mol Microbiol 23, 423-429.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228, 1315-1317.

Soelaiman, S., Jakes, K., Wu, N., Li, C., and Shoham, M. (2001). Crystal structure of colicin E3: implications for cell entry and ribosome inactivation. Mol Cell 8, 1053-1062.

Stengele, I., Bross, P., Garces, X., Giray, J., and Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. J Mol Biol 212, 143-149.

Sui, M. J., Tsai, L. C., Hsia, K. C., Doudeva, L. G., Ku, W. Y., Han, G. W., and Yuan, H. S. (2002). Metal ions and phosphate binding in the H-N-H motif: crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal ions. Protein Sci 11, 2947-2957.

Symmons, M. F., Welsh, L. C., Nave, C., Marvin, D. A., and Perham, R. N. (1995). Matching electrostatic charge between DNA and coat protein in filamentous bacteriophage. Fibre diffraction of charge-deletion mutants. J Mol Biol 245, 86-91.

Vetter, I. R., Parker, M. W., Tucker, A. D., Lakey, J. H., Pattus, F., and Tsernoglou, D. (1998). Crystal structure of a colicin N fragment suggests a model for toxicity. Structure 6, 863-874.

Webster, R. E. (1991). The tol gene products and the import of macromolecules into Escherichia coli. Mol Microbiol 5, 1005-1011.

Weiss, G. A., Wells, J. A., and Sidhu, S. S. (2000). Mutational analysis of the major coat protein of M13 identifies residues that control protein display. Protein Sci 9, 647-654.

Wiener, M., Freymann, D., Ghosh, P., and Stroud, R. M. (1997). Crystal structure of colicin Ia. Nature 385, 461-464.

Witty, M., Sanz, C., Shah, A., Grossmann, J. G., Mizuguchi, K., Perham, R. N., and Luisi, B. (2002). Structure of the periplasmic domain of Pseudomonas aeruginosa TolA: evidence for an evolutionary relationship with the TonB transporter protein. Embo J 21, 4207-4218.

Zakharov, S. D., and Cramer, W. A. (2002). Colicin crystal structures: pathways and mechanisms for colicin insertion into membranes. Biochim Biophys Acta 1565, 333-346.

Zhai, Y. F., Heijne, W., and Saier, M. H., Jr. (2003). Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor, MotAB. Biochim Biophys Acta 1614, 201-210.

SECTION 2
THE STRUCTURAL BASIS OF PHAGE DISPLAY

2.1 pIII and the phage infection mechanism

2.1.1 Introduction

Originally called adsorption protein (or protein A), pIII has been the main subject of attention of the first scientists working with filamentous bacteriophages. After the three different functional domains had been identified, different research groups tried to elucidate their different properties and were able to assign a particular function to each of the N1, N2 and C-terminal domains. A significant step was achieved with the publication of the NMR structure of the N1 domain alone (Holliger and Riechmann, 1996); the crystal structure of N1 in complex with N2 revealed a striking homology between the two domains, which may indicate they are evolutionarily related (Holliger et al., 1999; Lubkowski et al., 1998). Yet, the C-terminus still remains structurally uncharacterized, and its expression proved to be problematic (Beckmann et al., 1998; Martin and Schmid, 2003a; Rakonjac et al., 1999). Recently, Weiss et al. proposed a computational model of the three-dimensional structure of the C-terminus (Weiss et al., 2003).

It is always stated that between three and five copies of pIII are located at the end of the phage particle, and it has been proposed that they function as a multimer (Gailus et al., 1994), but pIII's have never been seen directly in electron microscopy. The presumed stoichiometry is based on the observation that a threshold pIII concentration of about 160 molecules per cell is needed for termination of phage assembly and formation of infectious particles. A Poisson calculation revealed that, with a low concentration of pIII in the cell, 200 assembly sites would be formed, which suggested that four or five molecules of pIII per assembly site are required for the formation of the infectious particle (Rakonjac and Model, 1998). As no indication of oligomerization of the N1-N2 complex are presented in the crystal structures, one can reasonably deduce that the multimeric arrangement of pIII is most likely caused by its C-terminal domain, or possibly its C-terminal hydrophobic anchor.

With this project, we wanted to investigate the biochemical and structural characteristics of the C-terminus of pIII by expressing it in large amounts in *E. coli.* The original idea of solving the crystal structure had to be abandoned due to the proteolytic processing of this fragment and its resulting instability in solution. Nevertheless, evidence is presented supporting the hypothesis that under certain circumstances pIII might have a trimeric arrangement *in vitro*.

2.1.2 Materials and Methods

Protein expression and purification - All constructs involving a fusion to protein D were designed as described in Figure 2-1. The sequences encoding nucleotide positions 1632-2852 (Long construct, i.e. with the C-terminal helix) or 1632-2768 (short construct, i.e. without the C-terminal helix) of pIII from phage cloning vector m13mp19 (Genbank accession # L08821) were polymerase chain reaction-amplified and cloned into the plasmid pAT37 (based on pQE30 from Qiagen), with an engineered enterokinase cleavage site, recognized by bovine enterokinase (purchased from Invitrogen). The different recombinant proteins were expressed and purified as specified in the text. Where explicitely mentioned in the text, denaturing conditions were used for protein purification. The protein was denatured overnight in a buffer containing 8 M urea and 50 mM DTT. After dialyzing out DTT overnight, the denatured protein was refolded by dialysis against 0.2 M Tris pH 9.0 (4°C), 2 mM EDTA, 0.5 mM benzamidine-HCl, 0.8 M arginine, 0.5 mM amino-n-caproic acid. After 24 hours refolding at 4°C under magnetic stirring, insoluble material was removed by centrifugation and the various constructs were processed differently. The constructs without pD fusion were dialyzed against MHA buffer (33 mM MES, 33 mM HEPES, 33 mM acetate, pH 8.0), and directly loaded on an anion exchange chromatography column (equilibrated at pH 8.0) on a BIOCAD 700 workstation (Applied Biosystems). The eluted fractions were collected and dialyzed against an appropriate buffer for further processing. The pD fusion constructs were digested with enterokinase at room temperature for 4 h in a buffer containing 50 mM Tris pH 8.0, 1 mM $CaCl_2$, 0.1% Tween-20, using 1 Unit per mg of fusion of the recombinant bovine enterokinase. The solution was then dialyzed against 100 mM MHA buffer pH 8.0. Removal of pD and enterokinase was achieved with the coupled IMAC/AIEX (anion exchange) protocol on the BIOCAD workstation, based on the different isoelectric points of pIII, pD and enterokinase.

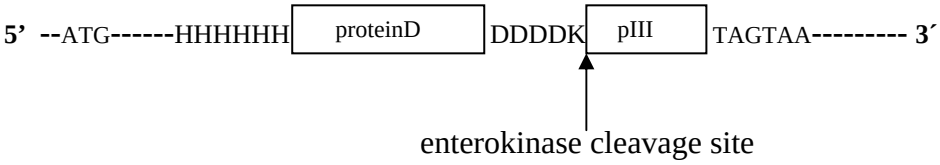


Figure 2-1. Design of the protein D fusions constructs, cloned into the vector pAT37, under the control of a T5 promoter (Forrer and Jaussi, 1998). Protein D is the major capsid protein of phage lambda; it contains an hexahistidine-tag at its N-terminus, and a smaller linker (GSGS) at the C-terminus. The recognition sequence for enterokinase (DDDK) is not originally present in the vector, and is engineered by PCR.

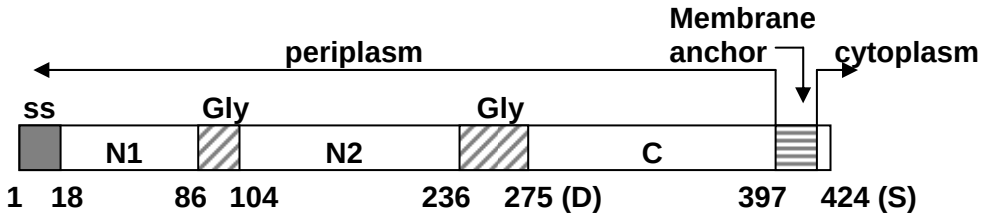
Molecular biology - DpnI site-directed mutagenesis was done as described in the specifications of the manufacturer (QuikChange Site-Directed Mutagenesis kit, Stratagene), in order to introduce the lysine (AAG) to alanine (GCG) mutation. The following primers were used: 5'-TCCGGTGATTTTGATTATGAAGCGATGGCAAACGCTAATA AGGGG-3' for the forward reaction, and 5'-CCCCTTATTAGCGTTTGCCA TCGCTTCATAATCAAAATCACCGGA-3' for the reverse reaction.
PCR conditions were as follows: 95°C (30 sec), and then 16 cycles of 95°C (30 sec), 55°C (1 min), 68°C (10 min) with 20-50ng of DNA template, 100 μM of each primer, 10x Pfu-polymerase buffer, 5 mM of each dNTPs, and 2.5 Units of Pfu polymerase. Mutation insertion was checked by DNA sequencing.

Western blotting – SDS-gel was blotted onto a PVDF membrane (Millipore, cat.# IPVH00010) previously soaked into methanol, and rinsed with water. After 1 h transfer, the membrane was blocked for 1 h with 5% milk/TBST, washed 3x with TBST, and incubated for 1 h with monoclonal antibody 10C3 (500x diluted in 2% milk/TBST). After 3 washing steps, the anti-mouse IgG-horseradish peroxydase secondary antibody (Pierce, cat.# 31438, 10'000x dilution) was added and further incubated for 1 h. After 5 sequential washing steps, development was done with 5 ml ECLA/ECLB 1 :1 mix (Amersham) during 1 minute. The detection was done on an X-ray film (Amersham)

2.1.3 Results

As we were interested in the C-terminal domain alone, we first tested the expression of the CT domain with and without the hydrophobic region (underlined in Figure 2-2b), thus giving rise to so-called pIII-Long and pIII-short, respectively (in both cases, without the signal sequence named “ss” in Figure 2-2a). The boundaries of the C-terminal anchor were determined on the basis of published data (Davis et al., 1985). Both constructs were cloned into the vector pTFT74 and expressed in *E. coli* as inclusion bodies in the strain BL21(DE3)pLysS (Studier, 1991). Both constructs gave no significant band on an SDS-PAGE gel; however, the pIII-Long construct could be detected on a Western blot (Figure 2-3). Other attempts in the strain C41(DE3) (Miroux and Walker, 1996), or BL21(DE3)pLysE gave no conclusive results. It might be that the membrane-active C-terminal fragment is lytic when cloned in such *E. coli* strains, enhancing the effect of the T7 lysozyme. Alteration of inner membrane permeability allows the passage of the lysozyme to the periplasmic space where it could exert its lytic activity on the bacterial outer membrane (Ciccaglione et al., 1998).

(a)



(b)

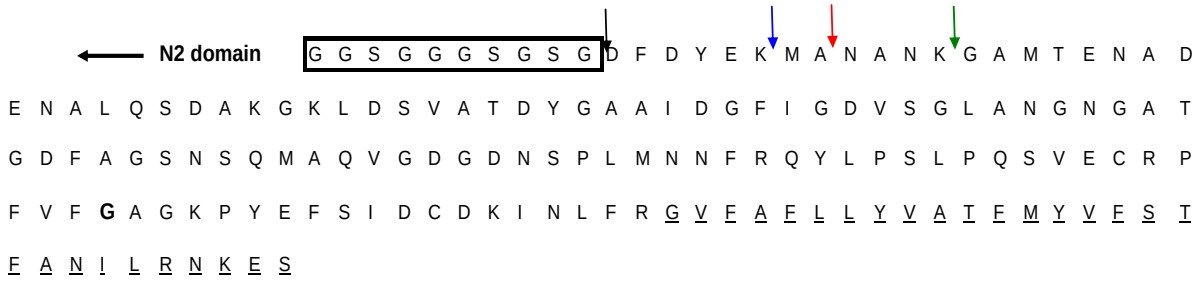


Figure 2-2. (a) General overview of pIII, with the three segmented domains (N1, N2 and C), delimited by the glycine-rich linkers. The 18 first amino acids contain the signal sequence (ss). The letters in parentheses (D/S) indicate the boundaries of the C terminus, whose sequence is detailed in (b). “Periplasm” and “cyto” mean the part of the protein which is located in the periplasm, or in the cytoplasm respectively, of *E. coli* before assembly. Taken from Rakonjac et al. (1999). (b) Amino acid sequence of the C-terminus ; the hydrophobic α-helical membrane anchor region is underlined. The end of the second glycine-rich linker is boxed. The desired cleavage site is indicated by a black arrow. The colored arrows show the unspecific cleavage sites, resulting from digestion with enterokinase or contaminating proteases from *E. coli* (see text for details). The residue in bold (4th position, 4th row) differs from the sequence in Swiss-Prot (entry PO3662), which has a serine at this position.

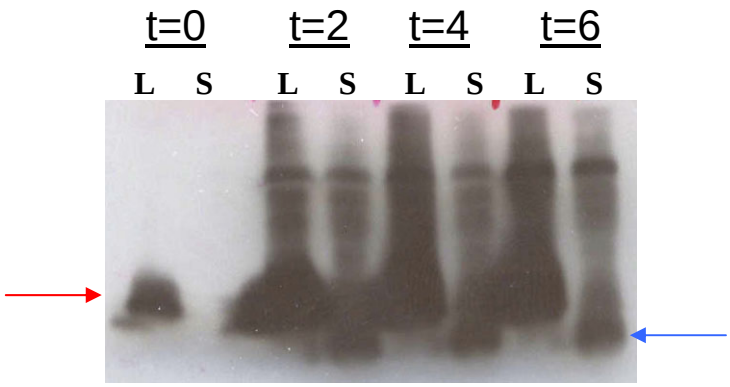


Figure 2-3. The long construct (L, red arrow) of the C-terminal domain of pIII is recognized by the monoclonal antibody 10C3 (Beckmann et al., 1998), it migrates on this blot at the proper molecular weight (14 kDa). The short construct (S, blue arrow) has a reduced expression yield, and is smaller (12 kDa). On top of each lane is indicated the time (in hours) after induction of protein expression. Whole cell extracts were boiled at 95°C for 10 minutes, loaded and run on a 15% SDS-gel, and blotted onto a PVDF membrane.

We fused the CT domain (again as long and short construct) with the protein D from bacteriophage Lambda, as this strategy meanwhile proved to be successful with TonB (see chapter 2.3). The cells reached an optical density of 1.4 (at 600 nm) after 2 hours expression at 30°C, then stopped growing before eventually dying, probably due to the toxicity of the protein. The fusion protein was expressed in soluble form in the *E. coli* strain XLI-Blue, and it showed an extensive degradation pattern when visualized on an SDS gel (Figure 2-4a). With the aid of a hexahistidine-tag at the N-terminus of protein D, both fusions were purified over nickel-NTA beads. Rapid processing of the protein thought to minimize the degradation propensity, did not prevent proteolysis (which, maybe, occurred already in the cell). The subsequent digestion with enterokinase in an effort to recover the pIII alone thus failed (Figure 2-4b). Because of the presence of two cysteines, most probably disulfide-bridged (Kremser and Rasched, 1994), we devised an alternative strategy to denature and fully reduce the protein in 8 M urea containing 50 mM DTT, dialyzing out the DTT and then refolding the fusion in a redox buffer at pH 9.0. Strong precipitation was observed upon removal of arginine (present in the refolding buffer) which would be necessary for subsequent purification over an ion-exchange chromatography column.

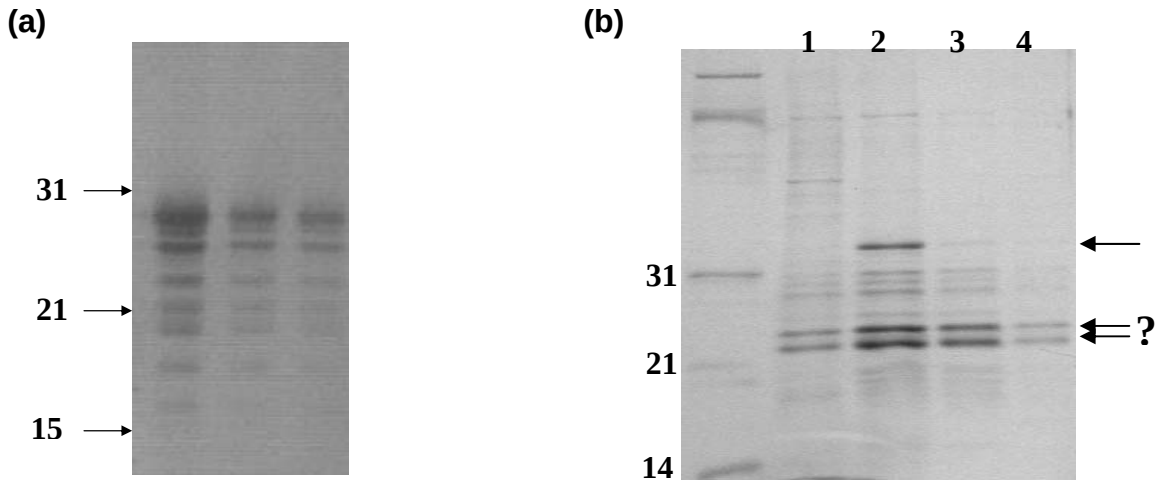


Figure 2-4. (A) SDS gel electrophoresis showing extensive proteolysis upon expression of pIII-short fused to protein D, and expressed in soluble from in *E. coli* (strain XL1-Blue). Three different clones of the same fusion construct are shown. The upper most band matches with the computed molecular weight (30 kDa). The numbers on the left side indicate the sizes in kilodalton. (B) batch purification of the pIII-Long construct (fused to protein D); the protein has been expressed (as described in *Material and Methods*) and loaded on a column containing nickel-NTA agarose beads. After washing (lane 1), the fusion protein is eluted with 0.5 M imidazol (upper arrow), along with degradation products (double-arrow with question mark). Both gels contain 15% SDS and are stained with Coomassie.

New results indicated that despite the presence of a long, flexible glycine-rich linker between all the domains of pIII, there is a specific interaction between N1-N2 and the CT domain, meaning that the latter cannot be regarded as an independent domain. This interaction was inhibited to varying degrees by the fusion of a peptide at the N-terminus of the CT domain (Chatellier et al., 1999).

Expression of the full pIII (containing N1, N2 and the CT domain cloned in pTFT74) was therefore tested, but no significant enrichment of the protein in inclusion bodies was observed when expressed in the strains BL21(DE3) or C41 (data not shown). We then devised a new strategy, involving again the fusion of the full pIII (N1, N2 and CT domain, without the signal sequence) to protein D. Again, two constructs were created with and without the hydrophobic anchor, thus giving rise to the so called “pD-pIII-Long” and “pD-pIII-short”, respectively. The absence of the C-terminal helix had a beneficial effect on the expression, and a thick band was clearly visible on the gel; pD-pIII-short was observed to be expressed in soluble form in XL1-Blue up to an OD(600nm) of 1.8 after 3 – 4 hours expression at 30°C, and no visible degradation occurred on the gel. The theoretical mass of the undigested pD-pIII-short fusion protein is 53’361 Dalton, which is considerably smaller than the band in lane 2 on Figure 2-5. However, due to the flexible glycine-rich linkers, pIII is known to display aberrant migration properties on SDS-gels (Crissman and Smith, 1984; Gray et al., 1981). The undigested fusion protein (lane 2, Figure 2-5) was subjected to mass spectrometry

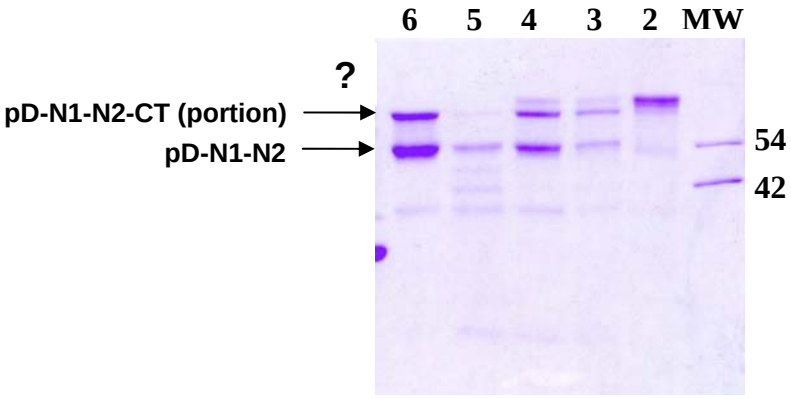


Figure 2-5. This 12% Coomassie-stained SDS gel illustrates the various steps of purification of the pD-pIII-short construct (53.3 kDa). Two molecular weight standards are indicated in the far right lane. After purification on a nickel column (IMAC) under native conditions, the fusion protein was denatured and reduced in 8 M urea (with 50 mM DTT). Dialyzing out the DTT allowed the refolding of the protein in a refolding buffer containing a redox shuffle (lane 2). The protein was then dialyzed against the enterokinase-buffer, in order to maximize the rate and the specificity of enterokinase cleavage, done for 4 h at RT (lane 3). The cleaved fusion protein was then loaded on an IMAC column coupled to an anion-exchange column (HQ), in order to quantitatively remove the protein D moiety, which is not visible on this gel. On lane 4 and 5 are loaded the flow-throughs of the IMAC column and the HQ column, respectively. Lane 6 contains the pure pIII, eluted from the HQ column (see text for more details).

analysis: the calculated mass of 52'700 Dalton was a first hint that some degradation might have occurred. Amino acid sequencing confirmed that the amino terminus of the protein was intact and, therefore, that the proteolysis started at the C-terminus. By mass computation, it could be easily deduced that the last four residues before the C-terminal helix (sequence "NLFR", see Figure 2-2) were actually missing.

The pure protein without the pD moiety specifically eluted from the HQ column (lane 6 in Figure 2-5), was subjected to mass spectrometry analysis, which gave a molecular weight of 41.2 kDa (data not shown). The theoretical mass of the pIII-short construct after digestion with enterokinase is 39.3 kDa. Unexpectedly, N-terminal amino acid sequencing revealed that our purified product started with the hexahistidine-tag sequence. Furthermore, the SDS-gel shown in Figure 2-5 was transfered onto a nitrocellulose membrane and the membrane was probed with an anti-tetra histidine-tag antibody. The two bands present on the gel in lane 6 were also visible on the blot, indicating that proteolysis did happen from the C-terminus of pIII-short (data not shown). Therefore, it was concluded that the enterokinase cleaved the fusion protein in an unspecific manner within pIII, at the first lysine residue in the C-terminal domain (see blue arrow in Figure 2-2). This lack of specificity is not a rare event and has already been documented (Choi et al., 2001). The first amino acid downstream of the recognition sequence of enterokinase was suggested to be important for specificity; in that respect, aspartic acid was proposed to be among the residues conferring the best cleavage specificity (Hosfield and Lu, 1999). The exact identity of the two bands in lane 6

(Figure 2-5) could not be resolved. However, mass spectrometry analysis and N-terminal amino acid sequencing of the sample loaded in lane 6 detected one single species, with a mass of 41.2 kDa and an intact hexa-histidine tag. It was concluded that pIII adopts two different conformers, displaying distinct migration properties on an SDS-gel. Alternatively, it is entirely plausible that the upper band in lane 6 corresponds to a molecule still containing a portion of the C-terminal domain. The lower band could then be the 41.2 kilodalton species where the proteolysis of the C-terminal domain has reached completion. However, it remains to be explained why only this lower band was detected by the mass spectrometry and the amino acid sequencing. The pD-pIII-Long construct behaved in the same way, except that the expression yield was lower (probably due to the lethality effect of the hydrophobic helix which tends to insert into the inner membrane).

The most evident experiment to perform next was to mutate the lysine residue after which the unspecific cleavage was thought to occur. Introduction of an alanine residue at this position, however, did not restore cleavage specificity. The protein was expressed and purified in the same way, as it was done for the above-mentioned construct. The same kind of investigations led us to the conclusion that the enterokinase (or a contaminating protease from *E. coli*) did cleave this time after the next occurring lysine residue in the sequence (see green arrow, Figure 2-2). Mass spectrometry analysis detected a species at 41.78 kilodalton, exactly matching with a product cleaved at this position.

We then came back to our original idea, namely the expression of the unfused, full length pIII (without the signal sequence) as inclusion bodies, under the control of the strong T7 promoter, now with a hexahistidine-tag at the N-terminus. Both constructs, with and without the C-terminal helix were processed in parallel. The long construct (i.e. with the C-terminal helix) could be expressed at some levels in BL21 (DE3), but was later found to form soluble aggregates resistant to detergents such as dodecyl-maltoside (0.5 mM) or CHAPS (8 mM), as confirmed by the presence of the protein in the void volume of a gel filtration chromatography (data not shown). The short construct (i.e. without the C-terminal helix) showed the best expression pattern in the strain BL21(DE3), albeit at lower levels than in fusion with protein D (Figure 2-6). It is this protein, termed pIIIfull-short, which was studied in greater detail.

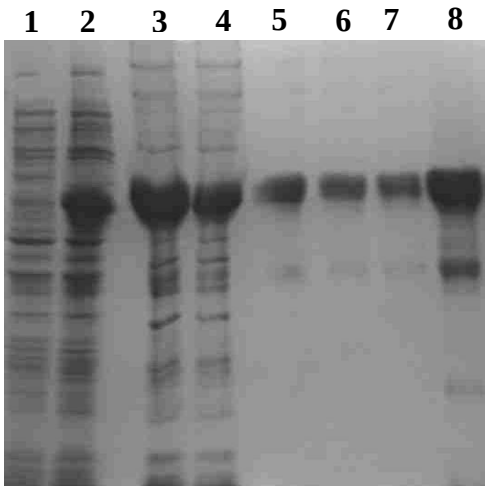


Figure 2-6. Expression and purification of pIII full-short (no fusion, no signal sequence, no C-terminal helix), inserted into pTFT74, for expression in inclusion bodies: (1) before induction, (2) crude extract after 3 hours induction at 37°C, the cells reached on OD_{600} of 2, (3) flow-through of IMAC purification under denaturing and reducing conditions (6 M GuHCl + 2 mM β-mercaptoethanol), (4) 2nd flowthrough, reloading of the first flow-through (lane 3) onto the column, (5) after refolding by dialysis in 0.8 M arginine, (6) after dialysis against a buffer suitable for anion-exchange chromatography, (7) elution with 700 mM NaCl from an anion-exchange column (BIOCAD) under native conditions, (8) same as lane 7, 10x concentrated. (15% SDS-PAGE, Coomassie-stained)

Curiously, quite a big proportion of pIIIfull-short did not bind to the IMAC column on the BIOCAD, even though this purification step was done under strong denaturing conditions and the disulfide bridges had been fully reduced by prolonged incubation with β-mercaptoethanol (see lane 3 and 4, on Figure 2-6). This phenomenon could not be circumvented, whatever the flow rate was when loading the extract onto the column; this may possibly reveal the presence of soluble aggregates. Again, the purified protein showed up as a double-band (lane 8), similar to what was observed with the protein D fusions (lane 6 in Figure 2-5). We were able to obtain 1 milligram of pure protein out of 1 liter culture; the protein was subjected to mass spectrometry, which detected a species at 40'362 Dalton, much in accordance with the computed mass of 40'363 Dalton (without the start Methionine).

A number of experiments were conducted in order to further characterize pIIIfull-short. The idea was to grow TG-1 F^+ cells (i.e. TG-1 cells harbouring the F-pilus) in the presence of varying concentrations of pIIIfull-short in the culture medium, and see if the bacteria might import it and were sensitive to the presence of the protein. Only a small quantity of pIII is known to cause strong membrane disturbances, which includes deoxycholate sensitivity, increased leakage of β–lactamase, enhanced tolerance to certain colicins, and defective F-pili (Boeke et al., 1982). Furthermore, it is well documented that *E. coli* cells induce the expression of the so-called "phage-shock-proteins" when challenged by phage proteins (Model et al., 1997). However, in our set-up, no specific accumulation of proteins of any kind could be detected following incubation with pIIIfull-short when compared to a control experiment with no protein (data not shown). The concentrated culture medium

was subjected to mass spectrometry, with the idea of seeing if any periplasmic *E. coli* protein (e.g. β-lactamase) could have leaked out of the cell following incubation with the protein. The molecular weight of the species present on the spectrogram did not fit to any *E. coli* protein, as told by the "TagIdent" software (available under http://au.expasy.org/tools/tagident.html). Similarly, the cells did not behave anomalously if detergent (e.g. Tween-20) was also present in the medium.

It was nevertheless decided to pursue the biochemical characterization of pIIIfull-short, and we next adressed the oligomeric state of the protein by loading it on a gel filtration column. We found it to be monomeric, in marked contrast to the pD-pIII-short construct thought to form trimers (Figure 2-7). A large scale purification was undertaken and 9 milligrams were sent to Alex Wlodawer's group for a first crystallization screening. An initial crystal survey yielded only clustered needle crystals of insufficient quality for further structural studies (Changsoo Chang, personal communication).

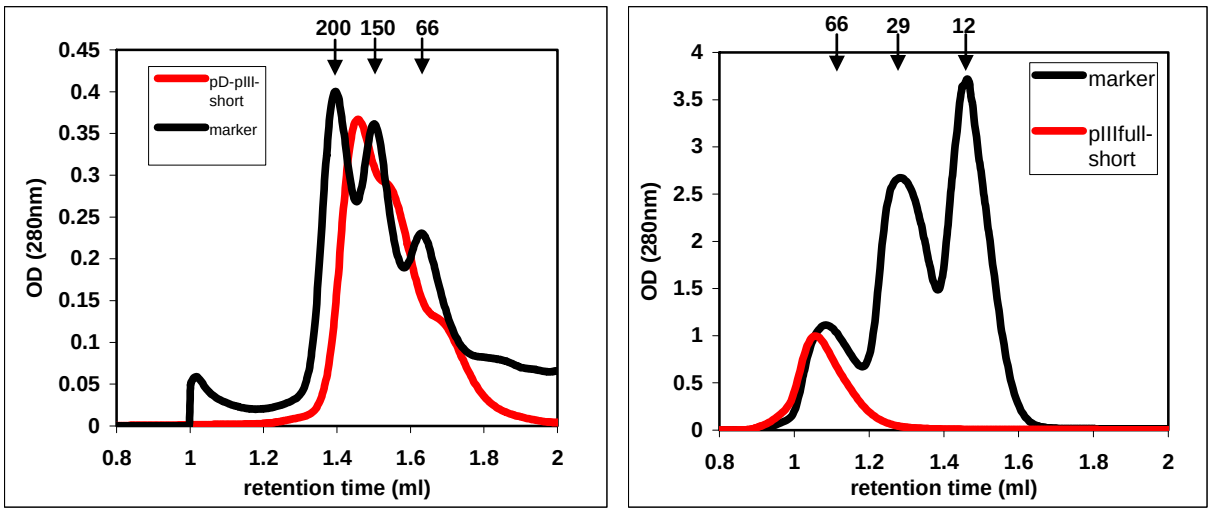


Figure 2-7. Elution profile of two different pIII constructs loaded on gel filtration chromatography columns. The left panel shows the behavior of the pD-pIII-short construct (in red) on a Superose 12 gel filtration column (SMART system, Pharmacia Biotech). The molecular weight standards are colored in black (200-150-66 kDa). The protein displays a main elution peak at ca.170-180 kDa, thus making it a trimer (the theoretical mass is 53'360 kDa). Even boiling at 95°C for 45 minutes does not affect significantly the elution behavior (a small amount of precipitate appeared). Protein D has been shown to be always monomeric on a gel filtration column (Patrik Forrer, personal communication).
On the right panel, pIIIfull-short construct (no fusion to protein D) has been loaded on a Superdex 75 column. The elution behavior (about 70 kDa) shows a huge discrepancy with the theoretical mass (40'363 kDa), probably due to the glycine-rich linkers causing a non-globular shape of the protein. From the molecular weight standards in black (66-29-12 kDa), one can deduce that pIIIfull-short is monomeric, and not aggregated as proved by running blue dextran (not shown).

In later investigations however, we did observe, to our great surprise, that pIIIfull-short was undergoing proteolysis after 2-3 days of storage at 4°C (or overnight at room temperature), as illustrated in Figure 2-8a. Mass spectrometry analysis (MALDI measurement) revealed that the molecular weight of this species was 28.4 kDa, matching with a sequence being truncated at the beginning of the C-terminal domain (see red arrow in Figure 2-2). The N-terminus was left intact, as confirmed by N-terminal amino acid sequencing. Basically, we ended up with the N1-N2 fragment (with the second glycine-rich region), whose structure had already been published (Holliger et al., 1999; Lubkowski et al., 1998). This fragment proved to be resistant to trypsin digestion. In order to test if the proteolysis was due to a contaminating protease carried along in the purification process (performed under strong denaturing conditions !), pIIIfull-short was incubated in the presence of a home-purified fusion protein (protein D-TonB). As pD-TonB was digested only in the presence of the truncated pIII, this could potentially mean that the contaminating protease should first attack pIII, before digesting the fusion pD-TonB. This observation led us to the conclusion that the putative contaminating protease should have a substrate preference for pIIIfull-short against proteinD-TonB (Figure 2-8b). The *E. coli* protease FtsH was immediately seen as a candidate responsible for this strange behavior, as its extreme resistance to denaturation has been documented (Cooper and Baneyx, 2001). As the activity of FtsH is known to be dependent upon ATP hydrolysis, ATP dependency was investigated, but no enhancement of the proteolysis through ATP was found (compare lanes 1 and 2 in Figure 2-8c).

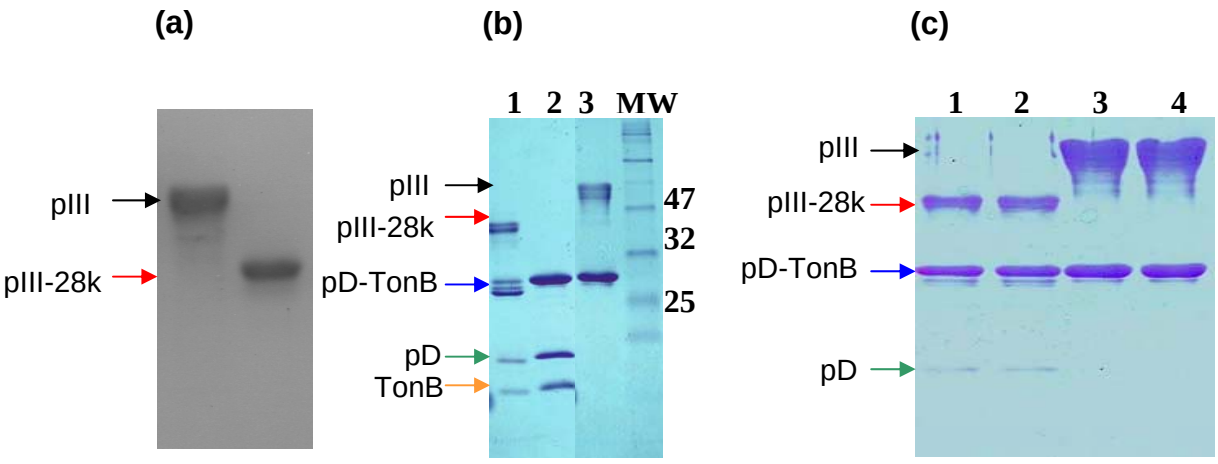


Figure 2-8. These Coomassie-stained SDS-gels illustrate the proteolysis by contaminating protease, which affects pIIIfull-short: (a) Some degradation already occurred in the freshly purified protein (left lane), becoming effective after overnight incubation at room temperature (right lane). (b) in lane 1, the truncated form (termed “pIII-28k”, red arrow) was mixed with pD-TonB (blue arrow, all three bands), and incubated overnight in Tris buffer, at room temperature: pD-TonB is partially digested into pD (green arrow) and TonB (yellow arrow). Lane 2 is a control where the pD-TonB fusion was incompletely digested with enterokinase. Curiously, no digestion is seen when the original unprocessed pIIIfull-short (black arrow) is mixed with pD-TonB (lane 3). The fourth lane contains the molecular weight standards. (c) “pIII-28k” (red arrow) was incubated with pD-TonB (lane 1), and supplemented with ATP (lane 2). A slight band corresponding to pD can be seen at the bottom of the gel (green arrow, confirmed by a Western blot not shown here). Similarly, pIIIfull-short and pD-TonB were incubated with and without ATP (lanes 3 and 4, respectively).

2.1.4 Discussion

In view of these results, it becomes evident that all attempts to purify the entire pIII protein have failed due to the impossibility to keep the full protein in solution in a stable form, and free of any fusion partner: the C-terminal domain gets proteolytically processed. The reason for the spontaneous proteolysis of the full-length pIII is still unresolved and remains an open question. All kinds of explanations can account for this behavior, but it still remains mysterious why the cleavage repeatedly occurs almost at the same position, no matter enterokinase is involved or not in the purification procedure. A contaminating protease, itself slowly refolding, may be one hypothesis. After it has refolded, it will cleave pIII, thus explaining why cleavage of pD-TonB is only seen in the presence of pIII-28k. Indeed, when looking at Figure 2-2, it becomes clear that the residues at the beginning of the C-terminus should have some propensity to form an unstructured and/or exposed region, particularly prone to proteolytic activity, despite the fact that structure prediction algorithms model this segment as an α-helix (data not shown). The susceptibility to proteolysis, as discussed in a recent review (Hubbard, 1998), is primarily determined by the ability of the protein to unfold locally, and this feature seems to be accomplished more easily in an α-helix than in a β-sheet. The rationale for this is that too many inter-strand hydrogen bonds would need to be broken and the ends brought closer together to deform an extended chain segment such as found in β-sheets. Interestingly, it has long been known that during its transfer from the phage coat to the inner membrane of the infected cell, pIII was reduced to a low molecular-weight form of about 13 kDa, based on its mobility in an SDS-gel (Marco et al., 1974). The question as to which part of pIII corresponds to the truncated form has never been adressed.

Furthermore, it must be stressed that maturational cleavage of phage proteins is not an unusual phenomenon. The attachment proteins of R17 and RNA phage Qβ are known to undergo such cleavage reaction upon attachment of the phage to the sides of the pilus (Manchak et al., 2002). In bacteriophage T4, a huge complex consisting of gene product 5 (gp5) and gp27 is required to penetrate the outer cell membrane of *Escherichia coli* and to disrupt the intermembrane peptidoglycan layer, thus allowing formation of a channel for injection of DNA into the host. Lysozyme activity, necessary to digest the peptidoglycan layer, is attributed to the lysozyme domain of gp5. When incorporated into the phage or stored at high concentrations, gp5 undergoes maturational cleavage resulting in two fragments which remain in the phage particle (Kanamaru et al., 2002).

The question of the stoichiometry of pIII in the phage coat still remains controversial. Indeed we obtained very contrasting results: pD-pIIIfull short did migrate as a trimer on a size-exclusion chromatography column, whereas the same protein not fused to protein D gave a monomeric peak, though at a higher size than expected (70 kDa, instead of 40 kDa, see Figure 2-7). However, as the latter proved to be prone to proteolysis, the results obtained with the undigested pD-pIIIfull-short construct look more reliable. In this respect, one can reasonably propose that pIII is displayed at the tip of the phage as a multimer (most probably trimer), and at least part of the C-terminus serves as an oligomerization module. Our last construct (pIIIfull-short), which actually lost the full C-terminus upon spontaneous proteolysis and showed monomeric behavior (Figure 2-7 and 2-8a) comes in support of this

hypothesis. It remains to be determined if oligomerization may have some physiological significance and to which extent it might affect the infection process.

Oligomerization of a protein for its proper membrane insertion is already a well documented phenomenon. The study of α-toxin from *Staphylococcus aureus* is an instructive case, giving some interesting insight into the mechanism of protein insertion into membranes. This toxin belongs to a family of pore-forming toxins that are essentially built of β-sheets and are thought to form porin-like channels. In order to form a transmembrane channel, the toxin has to undergo oligomerization and membrane insertion, a process shown to be powered by acidic pH. The protein is secreted as a soluble monomeric protein, and its N- and C- termini are separated by a protease-sensitive, glycine-rich loop. The N-terminus is the less pH-sensitive region of α-toxin, in contrast to the C-terminus, which has been proposed to unfold partially in a two-step process upon acidification (Vecsey-Semjen et al., 1996). At pH 3.5, it seems that the N-terminus is still fully folded, whereas the rest of the protein shows typical characteristics of the molten globule-like state necessary for membrane insertion to occur. The protein is sensitive to the local pH at the membrane surface, and this pH has to be lower than 4 for channel formation. Unfortunately, working with pIIIfull-short through a pH range of 5 to 10 did not diminish the extent of proteolysis, as visualized on an SDS-gel (data not shown).

The proteolysis of all our constructs is a critical issue which has to be adressed. Clearly, one has to find alternative ways to obtain a stable protein in solution, necessary for structural characterization or biophysical studies. If one assumes that the extensive degradation pattern of the domain is due to disordered structure resulting in protease sensitivity, it might be that the C-terminal domain may become folded upon interaction with its cognate partner. In this respect, pVI might be a good candidate for such an induced-folding mechanism, and its co-expression (under the same operon) might stabilize the C-terminus of pIII. It is well known that pVI by itself is highly hydrophobic, thus preventing any simple purification procedure, but it is not unreasonable that its association with pIII might shield some hydrophobic regions. The N1 and N2 domains provide the best example in support of this idea: N2 is known to precipitate heavily when expressed alone, whereas the N1-N2 complex can be expressed in soluble form and shows unusual stability properties (Martin and Schmid, 2003b). Another potential “folding-inducer” of pIII could be DNA, as both are known to interact with each other (Jazwinski et al., 1973).

Regarding our project, a contaminating protease is at the moment the single hypothesis which can account for the unusual degradation pattern of all our constructs. However, in this “contamination hypothesis”, it remains to be explained why this contamination first attacks pIII and then the fusion proteinD-TonB (see Figure 2-8b). It might be that pIII is more unstructured, thereby being a better substrate for the protease. Furthermore, it is quite odd that all our constructs are cleaved always at the same position (compare red, blue and green arrows in Figure 2-2). In other words, one can not totally exclude that a contamination has been carried all along the purification procedure, and that this contaminant is fairly resistant to harsh denaturing and reducing conditions (8 M urea and 50 mM DTT). FtsH could be a candidate for this action, as its metalloprotease active site remains functional in presence of 3-5M urea (Cooper and Baneyx, 2001). Expression of the

protein in an FtsH deficient strain could be a way to solve this problem (Qu et al., 1996). Therefore, for all the reasons mentioned, and as long as one can not prove the absence of any contamination, we still believe that the C-terminal domain of pIII is proteolysis-sensitive and cannot be obtained in a stable soluble form if expressed alone.

2.1.5 Literature

Beckmann, C., Haase, B., Timmis, K. N., and Tesar, M. (1998). Multifunctional g3p-peptide tag for current phage display systems. J Immunol Methods 212, 131-138.

Boeke, J. D., Model, P., and Zinder, N. D. (1982). Effects of bacteriophage f1 gene III protein on the host cell membrane. Mol Gen Genet 186, 185-192.

Chatellier, J., Hartley, O., Griffiths, A. D., Fersht, A. R., Winter, G., and Riechmann, L. (1999). Interdomain interactions within the gene 3 protein of filamentous phage. FEBS Lett 463, 371-374.

Choi, S. I., Song, H. W., Moon, J. W., and Seong, B. L. (2001). Recombinant enterokinase light chain with affinity tag: expression from Saccharomyces cerevisiae and its utilities in fusion protein technology. Biotechnol Bioeng 75, 718-724.

Ciccaglione, A. R., Marcantonio, C., Costantino, A., Equestre, M., Geraci, A., and Rapicetta, M. (1998). Hepatitis C virus E1 protein induces modification of membrane permeability in E. coli cells. Virology 250, 1-8.

Cooper, K. W., and Baneyx, F. (2001). Escherichia coli FtsH (HflB) degrades a membrane-associated TolAI-II- beta-lactamase fusion protein under highly denaturing conditions. Protein Expr Purif 21, 323-332.

Crissman, J. W., and Smith, G. P. (1984). Gene-III protein of filamentous phages: evidence for a carboxyl- terminal domain with a role in morphogenesis. Virology 132, 445-455.

Davis, N. G., Boeke, J. D., and Model, P. (1985). Fine structure of a membrane anchor domain. J Mol Biol 181, 111-121.

Forrer, P., and Jaussi, R. (1998). High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage lambda head protein D. Gene 224, 45-52.

Gailus, V., Ramsperger, U., Johner, C., Kramer, H., and Rasched, I. (1994). The role of the adsorption complex in the termination of filamentous phage assembly. Res Microbiol 145, 699-709.

Gray, C. W., Brown, R. S., and Marvin, D. A. (1981). Adsorption complex of filamentous fd virus. J Mol Biol 146, 621-627.

Holliger, P., and Riechmann, L. (1996). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5, 265-275.

Holliger, P., Riechmann, L., and Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 A: evidence for conformational lability. J Mol Biol 288, 649-657.

Hosfield, T., and Lu, Q. (1999). Influence of the amino acid residue downstream of (Asp)4Lys on enterokinase cleavage of a fusion protein. Anal Biochem 269, 10-16.

Hubbard, S. J. (1998). The structural aspects of limited proteolysis of native proteins. Biochim Biophys Acta 1382, 191-206.

Jazwinski, S. M., Marco, R., and Kornberg, A. (1973). A coat protein of the bacteriophage M13 virion participates in membrane- oriented synthesis of DNA. Proc Natl Acad Sci USA 70, 205-209.

Kanamaru, S., Leiman, P. G., Kostyuchenko, V. A., Chipman, P. R., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2002). Structure of the cell-puncturing device of bacteriophage T4. Nature 415, 553-557.

Kremser, A., and Rasched, I. (1994). The adsorption protein of filamentous phage fd: assignment of its disulfide bridges and identification of the domain incorporated in the coat. Biochemistry 33, 13954-13958.

Lubkowski, J., Hennecke, F., Pluckthun, A., and Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nat Struct Biol 5, 140-147.

Manchak, J., Anthony, K. G., and Frost, L. S. (2002). Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer. Mol Microbiol 43, 195-205.

Marco, R., Jazwinski, S. M., and Kornberg, A. (1974). Binding, eclipse, and penetration of the filamentous bacteriophage M13 in intact and disrupted cells. Virology 62, 209-223.

Martin, A., and Schmid, F. X. (2003a). Evolutionary stabilization of the gene-3-protein of phage fd reveals the principles that govern the thermodynamic stability of two-domain proteins. J Mol Biol 328, 863-875.

Martin, A., and Schmid, F. X. (2003b). A proline switch controls folding and domain interactions in the gene-3-protein of the filamentous phage fd. J Mol Biol 331, 1131-1140.

Miroux, B., and Walker, J. E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol 260, 289-298.

Model, P., Jovanovic, G., and Dworkin, J. (1997). The Escherichia coli phage-shock-protein (psp) operon. Mol Microbiol 24, 255-261.

Qu, J. N., Makino, S. I., Adachi, H., Koyama, Y., Akiyama, Y., Ito, K., Tomoyasu, T., Ogura, T., and Matsuzawa, H. (1996). The tolZ gene of Escherichia coli is identified as the ftsH gene. J Bacteriol 178, 3457-3461.

Rakonjac, J., Feng, J., and Model, P. (1999). Filamentous phage are released from the bacterial membrane by a two- step mechanism involving a short C-terminal fragment of pIII. J Mol Biol 289, 1253-1265.

Rakonjac, J., and Model, P. (1998). Roles of pIII in filamentous phage assembly. J Mol Biol 282, 25-41.

Studier, F. W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol 219, 37-44.

Vecsey-Semjen, B., Mollby, R., and van der Goot, F. G. (1996). Partial C-terminal unfolding is required for channel formation by staphylococcal alpha-toxin. J Biol Chem 271, 8655-8660.

Weiss, G. A., Roth, T. A., Baldi, P. F., and Sidhu, S. S. (2003). Comprehensive Mutagenesis of the C-terminal Domain of the M13 Gene-3 Minor Coat Protein: The Requirements for Assembly into the Bacteriophage Particle. J Mol Biol 332, 777-782.

2.2 Biochemical characterization of the C-terminus of TolR

2.2.1 Introduction

In line with the previous chapter, this second chapter deals with the phage infection process, but now from the bacterial “perspective”. The TolQRA pathway is studied in greater detail; and one if its component, the TolR protein, is the center of our investigation. The role provided by TolR during the phage infection mechanism still remains elusive. Its various interactions with the transmembrane helices of the neighbouring TolQ and TolA have been extensively described, but no precise function has ever been ascribed to TolR. This project was initiated in order to get a better understanding of the biochemical and biophysical properties of TolR, and possibly come to a crystal structure of its periplasmic C-terminal domain. It was in keeping with the general goal of better characterizing components of the TolQRA complex, as examplified by the crystal structure of the N1 domain with TolAIII (Lubkowski et al., 1999). From the beginning we were challenged by another group, which reported having obtained diffracting crystals of (almost) the same portion of TolR (Abergel et al., 2001). However, this structure has not been reported

Our data present evidence that the periplasmic portion of TolR (amino acids 41-142) displays high conformational flexibility, thus almost precluding any structural characterization. Elution properties of TolR on size exclusion chromatography reinforce the hypothesis that this portion of the protein is elongated in solution, thus confirming assumptions made by others. NMR spectra performed with two slightly different segments of TolR did not show a convincing dispersion pattern. Additionally, we present evidence that the portion of TolR from residues 54 to 142 is monomeric in solution.

2.2.2 Materials and Methods

- *Molecular biology*- The periplasmic part of TolR (residues 41–142, Swiss-Prot entry: P05829) has been fused to protein D and cloned into the vector pAT37 (see Figure 2-1, previous chapter) using the restriction sites *KpnI/HindIII*. The *tolR* gene was amplified from genomic DNA using the following primers (the enterokinase cleavage site is underlined):

5'-GGCCTTGGTACCGATGACGATGACAAGCAGAGCGTGGAGGTCGATC-3' (forward)
5'-CTATCAAGCTTATCAGATAGGCTGCGTCATTAAACC-3' (backward)

PCR reactions conditions were as follows: 92°C (2'); then 30 cycles at 92°C (30''); 58°C (45''); 72°C (45''); and a final elongation step of 10' at 72°C.

For the shorter constructs (residues 54-142), the backward primer was the same as above. Two shorter constructs were made: one cloned in pAT37 in fusion with protein D, using the following forward primer:

5'- GGCCTTGGTACCGATGACGATGACAAGCAGGCGGTGAGCAGTAACG-3'

and another construct, without protein D fusion, cloned into pQE30 (from Qiagen) using the following forward primer (underlined is an alanine spacer):

5'- GGCCTTGGATCCGCAGCGGCACAGGCGGTGAGCAGTAACG-3'

PCR reactions conditions were in both cases as described above.
After purifying the PCR products with the QiaQuick PCR purification kit (Qiagen), the DNA was digested by the restriction enzymes *KpnI/HindIII*, before being ligated into the similarly digested vector pAT37 for the pD constructs. The construct without pD fusion was digested with *BamHI/KpnI* and ligated into pQE30 (which contains a His-tag at the N-terminus, vector commercially available at Qiagen). Transformation into chemically competent *E. coli* XL1-Blue cells allowed one to identify single clones containing the correctly inserted, in frame *tolR* gene (checked by DNA sequencing)

-Protein expression and purification - Expression was done in LB medium in XL1-Blue cells as follows: from freshly transformed colonies on a plate, a single clone was picked and used to inoculate a culture containing 1% glucose and 100 μg/ml ampicillin. After overnight growth at 30°C (OD_{600} = 2.4), the culture was diluted 100x into fresh LB medium (supplemented with 100 μg/ml ampicillin, but without glucose) and allowed to grow until an OD_{600} of 0.8 was reached. Protein expression was induced with 1 mM IPTG and done overnight at 30°C. After overnight growth, cells usually reached an OD_{600} of 2.4. Cells were then harvested, resuspended in 100 mM Hepes buffer pH 8.0 containing 500 mM NaCl and 10 μg/ml DNAse. Cell lysis was performed with a French Press (Emulsiflex, Avestin). After centrifugation (30 minutes at 20'000 rpm), the supernatant was kept, filtrated (0.22 μm filter) and loaded onto a nickel column (IMAC) on a BIOCAD 700 perfusion chromatography Workstation (Applied Biosystems). TolR without fusion to protein D was then loaded onto a coupled anion exchange column (HQ), the elutate was collected and used for experiments. The pD fusion construct, eluted from the IMAC column, was dialyzed overnight against a buffer containing 50 mM Tris pH 8.0 and 2 mM $CaCl_2$; it was digested with enterokinase for 6 hours at room temperature, using 5 U/mg fusion. The last purification step involved a coupled IMAC/HQ (anion exchange) column on the BIOCAD, at pH 7.0. The flowthrough of

the HQ column was collected (!) as the TolR fragment did not bind to the matrix.

All the procedures involving the anti-protein D antibody fused to the chitin-binding domain were done as described (Blank et al., 2002).

Expression of the ^{15}N-labeled protein for NMR studies was done in the same way as described above, except that instead of LB, minimal M9 medium with the following composition was used: 10.6 g/L K_2HPO_4, 4.9 g/L KH_2PO_4, 0.5 g/L NaCl, 0.5 g/L $^{15}NH_4Cl$, 10 g/L glucose, 6.7 ml/L vitamin solution (purchased by Sigma, cat.# K-3129), 1 ml/L trace metal mix, 2 mM $MgSO_4$, 100 μM $CaCl_2$.

The overnight preculture was washed 3 times in M9 medium before being inoculated into fresh M9 medium for expression.

Light scattering study was performed on a Mini Dawn Tristar (Wyatt Technology) with the protein at 130 μM in a buffer containing 50 mM phosphate pH7.0, 150 mM NaCl, 0.005% Tween-20.

2.2.3 Results

As a first construct, the periplasmic portion of TolR covering residues 41 to 142 was investigated. As the strategy of expression with protein D proved to be successful with TonB (see next chapter), the same concept was applied here (Figure 2-9). Good expression yields for the fusion protein were obtained, which was already quite pure after a single purification step over an IMAC column (Figure 2-10). The fusion construct was subjected to digestion by enterokinase, and the products of the reaction were loaded on coupled IMAC / AIEX (anion exchange) chromatography columns to separate protein D, enterokinase and the TolR fragment based on their different isoelectric points. Enterokinase cleavage for quantitative removal of protein D proved to be rather unspecific, resulting in a protein D fragment cleaved from its N-terminal histidine-tag, thus preventing any additional purification step by IMAC (data not shown).

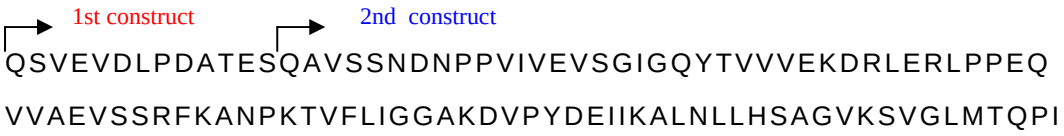


Figure 2-9. Amino acid sequence of the two constructs designed for the study of the periplasmic domain of TolR. The first one (red) starts with amino acid 41, whereas the second one (blue) starts at position 54 (numbering according to Swiss Prot entry P05829).

This impurity had to be removed by an alternative strategy. An anti-protein D antibody (from Morphosys, Martinsried, Germany) was genetically fused to a chitin-binding domain allowing specific binding of this fusion protein to a chitin column (Blank et al., 2002). By loading the product of the enterokinase digestion onto this affinity column, quantitative removal of the cleaved protein D fragment could be achieved. This purified sample of the C-terminal domain of TolR (lane 4, Figure 2-10) displayed a dimer/trimer equilibrium on a size exclusion chromatography, and circular dichroism attested the presence of some secondary structure, thus excluding any random coil conformation (Figure 2-11). Algorithms predicted an α–helical content of 13.7%, whereas the rest is composed of 49% β–sheet and 37.3% random coil.

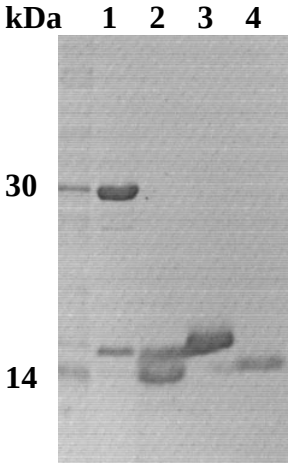


Figure 2-10. Expression and purification of the fusion proteinD-TolR, as visualized on a 17% SDS gel, loaded with the undigested fusion proteinD-TolR after IMAC purification (lane 1), fusion protein after cleavage with enterokinase (lane 2), protein D removed via IMAC purification (lane 3) and TolR, C-terminus (lane 4). The yield of the undigested fusion (lane 1) was 20 mg/L culture, and the final yield of the TolR C-terminus (lane 4) was 4 mg/L culture. The protein D fragment resulting from unspecific enterokinase cleavage is not visible here. Molecular weight markers are shown on the left.

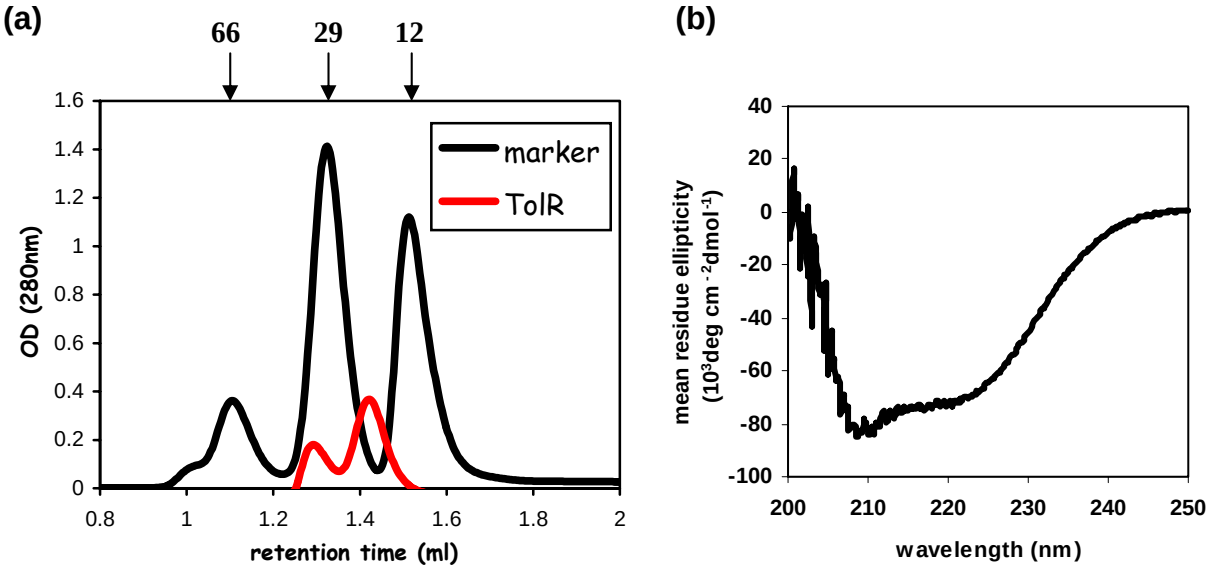
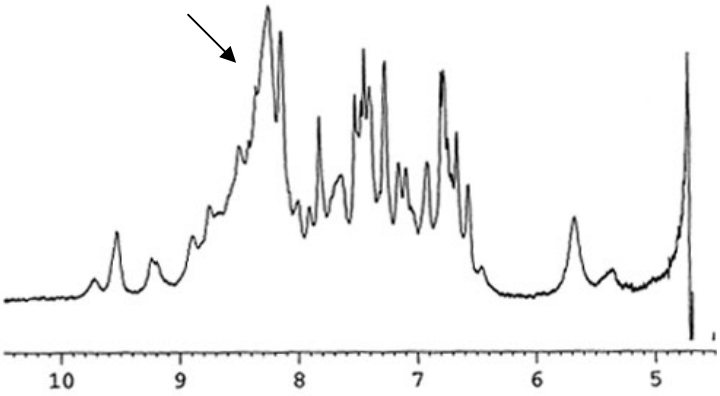


Figure 2-11. (a) TolR C-terminus (aa 41-142, MW = 10'940 Dalton), at 0.2 mg/ml in TBS_{150} pH 8.0, was loaded on a Superdex 75 gel filtration column (Pharmacia Biotech). TolR is eluted in two peaks of 24 and 35 kilodalton, suggesting the presence of a dimer/trimer equilibrium. As standards, BSA (66 kiloDalton), carbonic anhydrase (29 kDa) and cytochrome c (12 kDa) have been used. (b) the circular dichroism spectrum has been recorded with the protein in a 20 mM phosphate buffer pH 8.0, at a protein concentration of 0.74 mg/ml, in a 1 mm cuvette.

Mass spectrometry detected a main species at 10'939 Dalton (computed mass: 10'940 Dalton), and a side product of 9'569 Dalton, which is not visible on SDS-gels. Theoretical mass calculations predicted that this side product could be a portion of TolR, with its first thirteen amino acids deleted. This N-terminal portion of TolR might adopt conformational flexibility, rendering it sensitive to any protease contaminant. In view of the deceiving results obtained in the crystallization trials (formation of aggregated needle-shaped crystals), it was decided to subject the protein to limited proteolysis through incubation with the V8 protease from *Staphylococcus aureus*, as this strategy might sometimes be successful in yielding a crystallizable construct (Chen et al., 1998). Unfortunately, this treatment did not affect the crystallization properties of our protein either (Changsoo Chang, personal communication)

We then reasoned that NMR might possibly be more suitable to the study of TolR, showing such a high content of conformational flexibility. Thus, a one-dimensional spectrum was recorded at pH 6.0, which shows a rather good dispersity (Figure 2-12a). An ^{15}N-labeled sample was later expressed and purified for in depth analysis, but this preparation proved to be highly conformationally inhomogenous (Figure 2-12b). Based on this result, it was decided to remove the putative flexible part and a new construct was engineered, starting from amino acid residue 54 (Figure 2-9, blue), but now expressed without any protein D fusion. HSQC spectra run with an ^{15}N-labeled protein look very similar to the one presented in Figure 2-13a. This result was quite surprising, as this protein (TolR 54-142, ^{15}N-labeled) gave an almost ideal dispersion behavior in a light scattering measurement (Figure 2-13b). Interestingly, this last experiment detected the presence of a clear monomeric species in solution, which is worth being compared with other studies (Abergel et al., 2001; Journet et al., 1999). As these two publications were dealing with a fragment extending from residues 44 to 142, it is tempting to speculate that the 10 residues removed at the N-terminus of our construct might form a dimerization module.

(a)



(b)

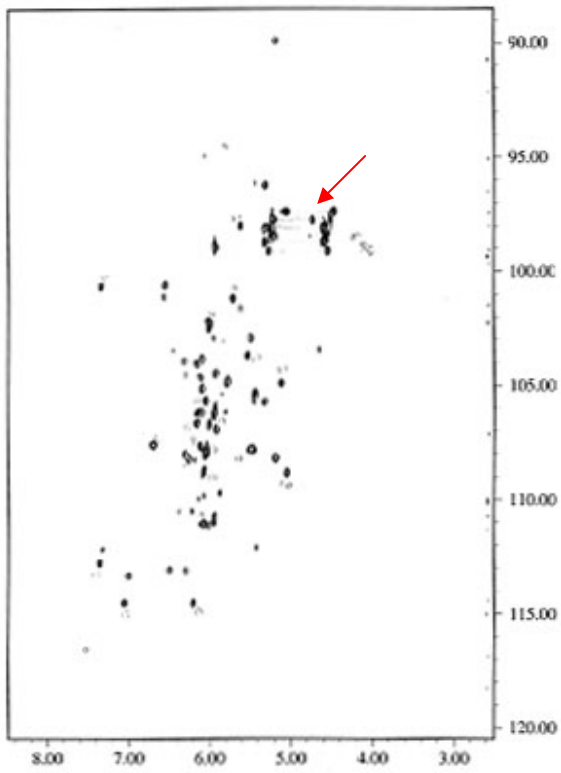
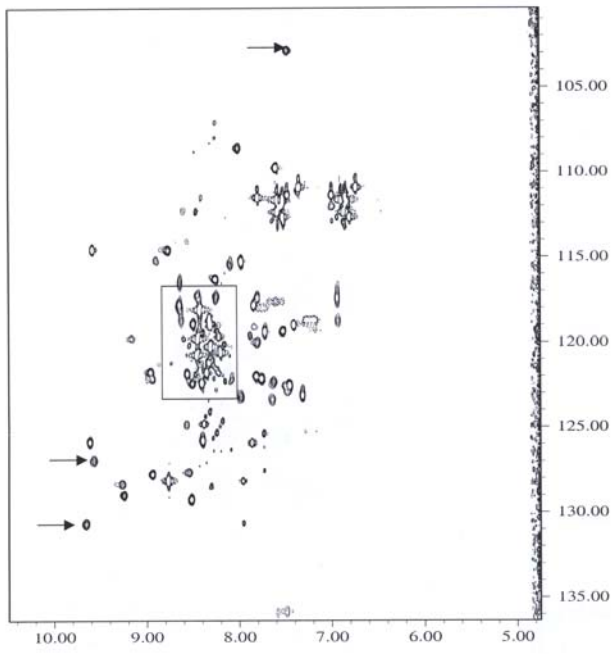


Figure 2-12. NMR spectra of TolR, C-terminal domain (amino acids 41-142).
(a) The interesting region (6-9 ppm) in this one-dimensional spectrum shows sharp peaks, indicative of good dispersion properties, as well as a flexible region at the N-terminus (arrow). (b) 2D spectrum of an ^{15}N-labeled sample showing high conformational inhomogeneity. Only 6 glutamine and asparagine residues are visible here (red arrow), whereas there should be nine, thus reinforcing the intuition that some proteolytic processing occurs at the flexible N-terminus. Some sharp peaks do show up, which are probably coming from contaminating protein D (Prof. Hideo Iwai, personal communication).
Both spectra were taken in 10 mM phosphate buffer pH 6.5, with 1 mM TolR.

(a)



(b)

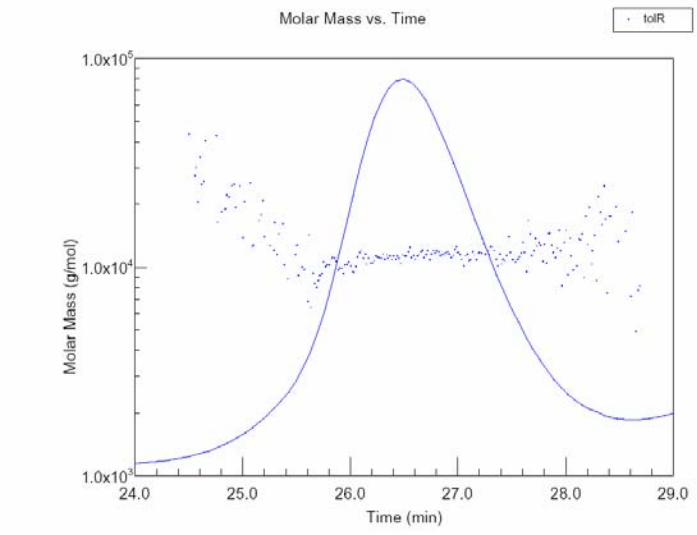


Figure 2-13. (a) HSQC spectrum (pH 5.0, 27°C) of the TolR fragment (amino acids 54-142). About half of the protein looks unfolded. A large overlap in one region is boxed; this did not change upon varying pH, temperature or salt conditions. The line width of the spread peaks (indicated by arrows) are broad for this molecular size, indicating the chemical exchanges. The other peaks are very sharp, which is typical for flexible tails. Different temperature and pH conditions were tested, but all spectra look similar (personal communication from Prof. Dr. Hideo Iwai) (b) light scattering pattern of this same construct of TolR, attesting the presence of a monomeric and homogenous protein (calculated molar mass moment = 11'330 g/mol).

2.2.4 Discussion

"No longer can binding be viewed as simply a "lock and key" event or as an interaction involving rigid macromolecular surfaces". This citation, taken from a review (Dyson and Wright, 2002), summarizes the whole problematic encountered with this project. Because protein functions are based on precise arrangements of atoms, there was generally a rather naive belief that well-defined and fairly rigid protein structures as revealed by X-ray crystallography are the norm. However, it has been recognized only recently that numerous proteins lack intrinsic globular structure or contain long disordered segments under physiological conditions, and furthermore, that this is their normal, functional state. Partial disorder of a molecule means that the conformation of that portion has flexibility as well as adaptabiity. In many macromolecular assemblies (e.g. virus or flagellum assembly), conformational adaptability is an essential requirement in forming intricate molecular interactions. The assembly of icosahedral viruses, such as simian virus 40 or polyoma virus, cannot occur with predetermined rigid structures of the coat protein subunits. The C-terminal chains must be disordered and flexible until they meet with the partner structures to which they adapt their conformation (Namba, 2001).
In bacteriophage λ, the FII protein (gpFII) is a structural protein responsible for the joining of phage heads and tails. Its NMR structure displayed two large unstructured regions at the N-terminus and a large loop near the middle of the protein. The authors speculated this unstructured region becomes structured upon assembly: gpFII is present in multiple copies in the phage and is believed to assemble into a stable multimeric structure only when the correct target surface provided by phage heads is contacted (Maxwell et al., 2002). Therefore, it is not totally excluded that, in the case of TolR, the folding of the C-terminus into a compact domain requires a partner structure. In its absence, the folding process is incomplete and results in partial disorder of the domain. The search for having the right partner to achieve folding of TolR is a rather risky guess, and some potential candidates include the C-terminus of TolA, the transmembrane helix of TolR itself or the third transmembrane helix of TolQ (see below). But sometimes the interacting partners are not present in the immediate vicinity of each other. In this respect, the case of colicin N reveals an interesting feature: its translocation domain is intrinsically disordered, and there are now pieces of evidence that upon binding the C-terminus of TolA, the disordered structure converts to a folded domain; the TolA receptor can be seen as a folding template allowing the colicin to find its final structured state (Anderluh et al., 2004). Finally, one of the most compelling suggestions for the participation of unstructured proteins in binding interactions was provided by Shoemaker *et al.* (Shoemaker et al., 2000). The authors envisage that an unstructured protein would have a greater "capture radius" for a specific binding site than the folded state with restricted conformational flexibility and smaller fluctuations. The image that they conjure up is of a "fly-casting" mechanism, whereby the unfolded state binds weakly at a large distance and then folds as it "reels in" its target, thereby completing folding and binding simultaneously.

Handling of proteins with high conformational flexibility and polydispersity prompted researchers to study these proteins using hyperthermophilic strains, as exemplified by the C-terminus of FliG, a component of the rotor in the bacterial flagellar motor. This portion of the *E. coli* protein failed to crystallize, however, the same domain from the hyperthermophile

Thermotoga maritima was successful in giving crystals, and a structure at 2.2 Å resolution was determined (Lloyd et al., 1999). Unfortunately, the only thermophilic strain known to possess the tol-pal gene cluster is *Methanobacterium thermoautotrophicum*, but the homology of the putative TolR equivalent is very weak. It is believed that the tolQ/R gene pair in this strain was the original starting point for the evolution of this system, and that it was subsequently embellished during the development of Gram negative bacteria (Sturgis, 2001). Alternatively, one could think of designing a new construct starting at position 66 (amino acids "VEVSG….", see Figure 2-9), as the 10 amino acids located upstream are predicted to be disordered, and the downstream sequence shows some globularity. These assumptions are based on a software, that allows the user to plot the tendency within a query protein for order/globularity and disorder (available at http://globplot.embl.de and further discussed in Linding et al. (2003)).

Our results support the hypothesis that a dimerization module might be located in the N-terminal region of our first construct, more precisely between the residues 41 and 54. Indeed, when removing these thirteen amino acids, the protein shows an unambiguous monomeric behavior (Figure 2-13b), whereas the first construct tends to form dimers or even trimers (Figure 2-11a). Interestingly, crosslinking studies performed in the analogous TonB system found that ExbB (TolQ homologue) and ExbD (TolR homologue) were forming homodimers and homotrimers in vivo (Higgs et al., 1998); these authors later established the precise stoichiometry of the complex, namely one TonB for two ExbD and seven ExbB (Higgs et al., 2002). The presence of a dimeric TolR has also been reported by others when performing *in vivo* crosslinking studies with a construct extending from position 44 to 142, a so-called "TolRII-III" (Journet et al., 1999). The question of the stoichiometry of TolR in the inner membrane of *E. coli* has recently been adressed in an alternative way, in view of the structural and functional homologies of the TolQRA system with MotA and MotB of the flagellar motor. Based only on homology modeling with the MotAB system, these authors proposed an inner membrane unit consisting of $TolQ_4TolR_2TolA_1$. Furthermore, despite any sequence homology between the C-terminus of TolR and MotB, they see the involvement of the C-terminus of TolR as a bilayer-penetrating loop, responsible for forming an ion-selectivity filter, as seen for MotB (Cascales et al., 2001). MotB disposes of a large amount of sequence (about 140 residues) between the C-terminus of the transmembrane α-helix and a periplasmic region, called peptidoglycan binding motif. These authors have proposed that this sequence might function as a linker between the MotA/MotB complex and the cell surface. This linker is divided into two parts, the first (about 90 amino acids) is wholly or party dispensable and the second is indispensable (Figure 2-14). They propose that the first part is a random coil with little elastic properties and the second part forms an ordered domain, whose compact structure allows it to function like a spring and provide suitable elastic coupling between the Mot complexes and the cell surface (Muramoto and Macnab, 1998). If these dynamic properties of the periplasmic segment of MotB has any relevance for TolR is still questionable; TolR and MotB show a consensus sequence only in their transmembrane domains. There is now a growing body of evidence that MotB forms a symmetric dimer via its transmembrane α-helix, and that this association is direct and does not require MotA (Braun and Blair, 2001).

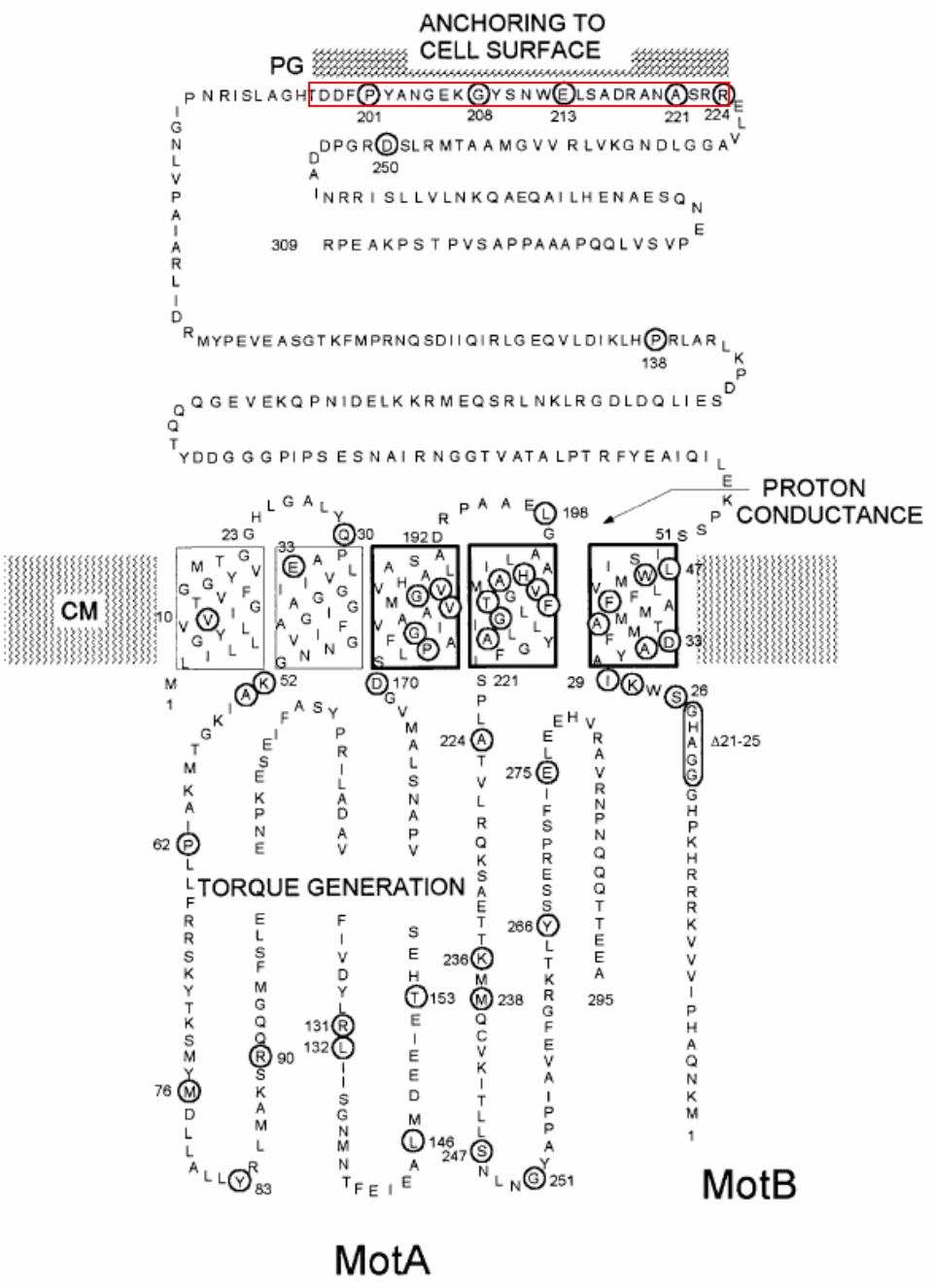


Figure 2-14. Cartoon of MotA and MotB showing their topology with respect to the cytoplasmic membrane (CM) and cytoplasmic and periplasmic regions. The transmembrane regions are presumed to be α-helical, but the boundaries of the transmembrane helices (boxes) should be regarded as only approximate. In MotB, the first linker region (position 51 to 138) thought to form a random coil, is dispensable for proper function. The second linker region lies in the vicinity of the peptidoglycan layer (PG). It forms a structurally ordered domain and is thought to be essential for the stability of MotB. Proposed roles of regions of sequence in proton conductance (heavily boxed transmembrane helices) and torque generation are indicated. The peptidoglycan binding motif necessary for anchoring to the cell surface is boxed in red. The circles denote positions of suppressor mutations. Picture taken from Muramoto and Macnab (1998).

The question of the natural interacting partners of TolR *in vivo* still remains a critical issue and forms a prerequisite for a structural characterization of its flexible C-terminus. The extreme C-terminus was suggested to interact either with the third transmembrane helix of TolQ and/or the transmembrane segment of TolR (Lazzaroni et al., 1995). Furthermore, these authors showed that TolRIII (residues 117 to 142) participates in the TolQ-TolR interaction; however, TolR deleted of this domain still crosslinked to TolQ. Therefore, the issue of knowing if co-expression of the C-terminal fragment with one or both of these transmembrane helices might have a stabilizing effect on the TolR conformation in solution remains highly hypothetical. Only the interactions between the transmembrane domains of TolQ, TolR and TolA have been well established and confirmed.

The *in vivo* localization of TolRII-III (residues 44 to 142) has been investigated; this portion of TolR appeared to be enriched in fractions with a sucrose density gradient intermediate between the fractions corresponding to the inner membrane and the outer membrane. This result argued in favor of an interaction of the TolR C-terminal domain with the membranes, but the authors went further by suggesting a localization of this domain in the contact sites between the inner and outer membranes (Journet et al., 1999). It was further demonstrated that the region from amino acid position 117 to 142 (so-called TolRIII) is not essential for TolR dimerization, and it plays a role in the TolA-TolR interaction, because deletion of this domain abolished the TolA-TolR complex. Journet et al. concluded from this finding that TolRIII plays a role in the TolA-TolR interaction, insisting on the fact that there exists no proof of a direct interaction between these two members, as no complex has ever been isolated or detected *in vivo*. Their results rather suggested that TolRIII interacts with proteins localized in the contact sites and/or the outer membrane, concordant with the results of Lazzaroni et al., who found that the C-terminus of TolR was not accessible to carboxypeptidase digestion (Lazzaroni et al., 1995). In the case of ExbD, a role for its C-terminal domain in the interaction with ExbB and TonB has also been suggested (Braun, 1995). It clearly appears then that the C-terminus of TolR is engaged in many interactions with neighbouring partners of the periplasm, and a more detailed understanding of the biochemistry of the TolQRA system is a prerequisite for structural characterizations of its members.

2.2.5 Literature

Abergel, C., Journet, L., Chenivesse, S., Gavioli, M., and Lloubes, R. (2001). Crystallization and preliminary crystallographic study of the periplasmic domain of the Escherichia coli TolR protein. Acta Crystallogr D Biol Crystallogr 57, 323-325.

Anderluh, G., Gokce, I., and Lakey, J. H. (2004). A natively unfolded toxin domain uses its receptor as a folding template. J Biol Chem 279, 22002-22009.

Blank, K., Lindner, P., Diefenbach, B., and Plückthun, A. (2002). Self-immobilizing recombinant antibody fragments for immunoaffinity chromatography: generic, parallel, and scalable protein purification. Protein Expr Purif 24, 313-322.

Braun, T. F., and Blair, D. F. (2001). Targeted disulfide cross-linking of the MotB protein of Escherichia coli: evidence for two H(+) channels in the stator Complex. Biochemistry 40, 13051-13059.

Braun, V. (1995). Energy-coupled transport and signal transduction through the gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. FEMS Microbiol Rev 16, 295-307.

Cascales, E., Lloubes, R., and Sturgis, J. N. (2001). The TolQ-TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA-MotB. Mol Microbiol 42, 795-807.

Chen, G. Q., Sun, Y., Jin, R., and Gouaux, E. (1998). Probing the ligand binding domain of the GluR2 receptor by proteolysis and deletion mutagenesis defines domain boundaries and yields a crystallizable construct. Protein Sci 7, 2623-2630.

Dyson, H. J., and Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 12, 54-60.

Higgs, P. I., Larsen, R. A., and Postle, K. (2002). Quantification of known components of the Escherichia coli TonB energy transduction system: TonB, ExbB, ExbD and FepA. Mol Microbiol 44, 271-281.

Higgs, P. I., Myers, P. S., and Postle, K. (1998). Interactions in the TonB-dependent energy transduction complex: ExbB and ExbD form homomultimers. J Bacteriol 180, 6031-6038.

Journet, L., Rigal, A., Lazdunski, C., and Benedetti, H. (1999). Role of TolR N-terminal, central, and C-terminal domains in dimerization and interaction with TolA and tolQ. J Bacteriol 181, 4476-4484.

Lazzaroni, J. C., Vianney, A., Popot, J. L., Benedetti, H., Samatey, F., Lazdunski, C., Portalier, R., and Geli, V. (1995). Transmembrane alpha-helix interactions are required for the functional assembly of the Escherichia coli Tol complex. J Mol Biol 246, 1-7.

Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31, 3701-3708.

Lloyd, S. A., Whitby, F. G., Blair, D. F., and Hill, C. P. (1999). Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor. Nature 400, 472-475.

Lubkowski, J., Hennecke, F., Plückthun, A., and Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure Fold Des 7, 711-722.

Maxwell, K. L., Yee, A. A., Arrowsmith, C. H., Gold, M., and Davidson, A. R. (2002). The solution structure of the bacteriophage lambda head-tail joining protein, gpFII. J Mol Biol 318, 1395-1404.

Muramoto, K., and Macnab, R. M. (1998). Deletion analysis of MotA and MotB, components of the force-generating unit in the flagellar motor of Salmonella. Mol Microbiol 29, 1191-1202.

Namba, K. (2001). Roles of partly unfolded conformations in macromolecular self-assembly. Genes Cells 6, 1-12.

Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 97, 8868-8873.

Sturgis, J. N. (2001). Organisation and evolution of the tol-pal gene cluster. J Mol Microbiol Biotechnol 3, 113-122.

Crystal Structure of the Dimeric C-terminal Domain of TonB Reveals a Novel Fold

Changsoo Chang[1#], Alexandre Mooser[2#], Andreas Plückthun[2*], and Alexander Wlodawer[1]

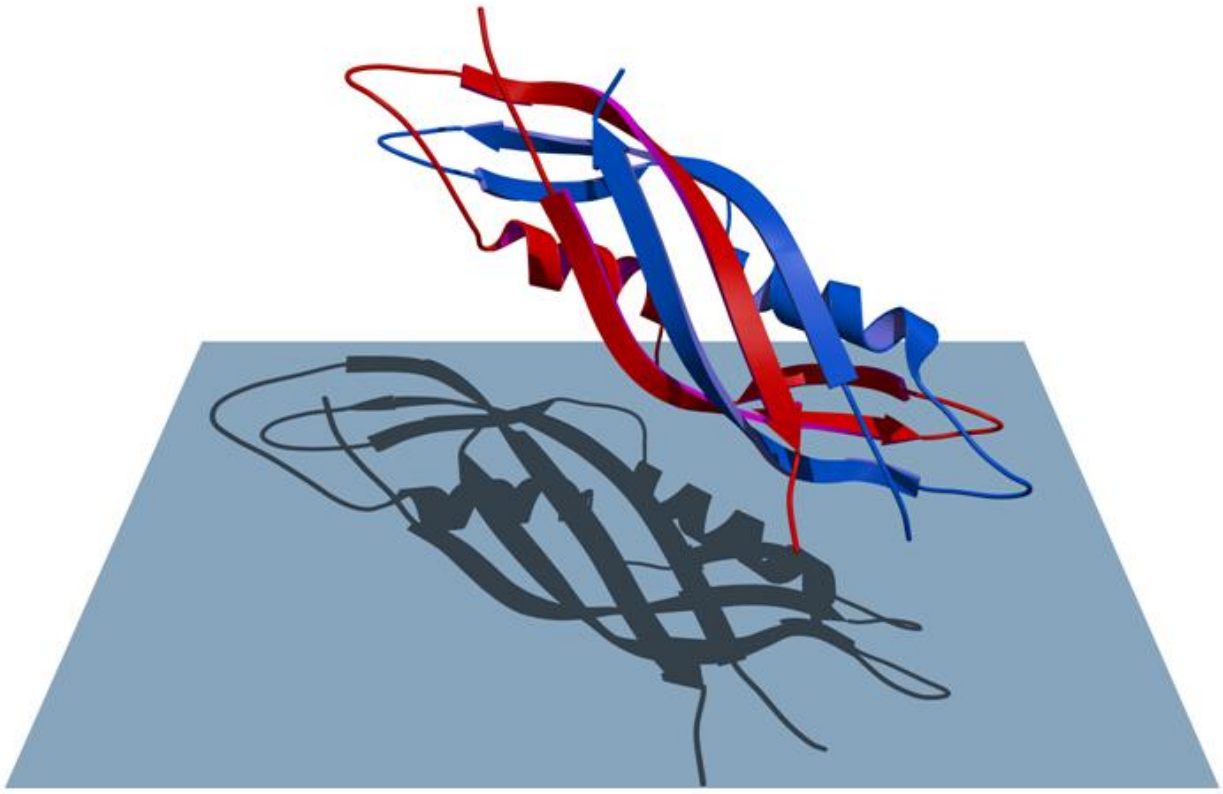
[1]Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD 21702, USA; [2]Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Running title: Crystal structure of TonB

[#]These authors contributed equally to this work

[*]Corresponding author:

Biochemisches Institut der Universität Zürich,
Winterthurerstrasse 190,
CH-8057 Zürich, Switzerland
Phone: (+41-1)-635-5570
Fax: (+41-1)-635-5712
E-mail: plueckthun@bioc.unizh.ch



THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 276, No. 29, Issue of July 20, pp. 27535–27540, 2001
Printed in U.S.A.

Crystal Structure of the Dimeric C-terminal Domain of TonB Reveals a Novel Fold*

Received for publication, March 29, 2001, and in revised form, April 26, 2001
Published, JBC Papers in Press, April 27, 2001, DOI 10.1074/jbc.M102778200

Changsoo Chang‡§, Alexandre Mooser§¶, Andreas Plückthun¶‖, and Alexander Wlodawer‡

From the ‡Macromolecular Crystallography Laboratory, NCI, National Institutes of Health, Frederick, Maryland 21702 and the ¶Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, Zürich CH-8057, Switzerland

The TonB-dependent complex of Gram-negative bacteria couples the inner membrane proton motive force to the active transport of iron·siderophore and vitamin B_{12} across the outer membrane. The structural basis of that process has not been described so far in full detail. The crystal structure of the C-terminal domain of TonB from *Escherichia coli* has now been solved by multi-wavelength anomalous diffraction and refined at 1.55-Å resolution, providing the first evidence that this region of TonB (residues 164–239) dimerizes. Moreover, the structure shows a novel architecture that has no structural homologs among any known proteins. The dimer of the C-terminal domain of TonB is cylinder-shaped with a length of 65 Å and a diameter of 25 Å. Each monomer contains three β strands and a single α helix. The two monomers are intertwined with each other, and all six β-strands of the dimer make a large antiparallel β-sheet. We propose a plausible model of binding of TonB to FhuA and FepA, two TonB-dependent outer-membrane receptors.

The outer membrane (OM[1]) of Gram-negative bacteria constitutes a permeability barrier, protecting the cell against a variety of toxic agents. The lipopolysaccharides located in the outer leaflet of the OM confer to the bacteria a polar and negatively charged surface, restricting the cellular uptake of toxic organic molecules and detergents such as bile salts, the detergents in the gut. However, although the OM is an effective protective barrier against harmful environmental components, it also represents an additional obstacle for the uptake of nutrients, which can be circumvented in three ways. While small hydrophilic nutrients (<600 Da) enter the periplasm by simple diffusion through porins in a non-selective manner (1), larger molecules are taken up by pores with an internal binding site for the ligand (such as LamB) in a stereospecific and selective manner (2) and can subsequently enter the cytoplasm by a variety of transporters located in the inner membrane (3). A few nutrients, notably iron and vitamin B_{12}, need to be taken up into the periplasm against their concentration gradients. For this purpose, a complex consisting of TonB, ExbB, and ExbD couples the inner membrane proton motive force (pmf) to the active transport of iron siderophores and vitamin B_{12} across the OM through specialized porins. Recently, crystal structures were solved for two TonB-dependent receptors, FepA and FhuA (4–6). Like all other known porins, they are β-barrels, but unlike the other porin structures they have much larger interiors, which are almost completely obscured by a protein domain sitting inside the barrel (termed the "cork" or "hatch region"), which is encoded within the N-terminal segment of either protein.

Iron uptake into bacteria is initiated by the binding of the iron·siderophore complex to the high affinity OM receptor. The dissociation constant is around 100–200 nM (7, 8). An electron spin resonance study (9), later rationalized by three-dimensional structural models (4–6), has shown that this event triggers conformational changes in the OM receptor. This might allow TonB to contact specific regions on the receptor. It appears that "energized" TonB is then able to deliver its energy to the receptor, resulting in ligand translocation into the periplasm (10, 11). ExbB·ExbD are implicated in the recycling of TonB, from its high affinity OM receptor association to a high affinity inner membrane association (12, 13). The structural changes in this whole process have remained almost completely unclear.

TonB of *Escherichia coli* is a protein consisting of 239 amino acids. Homologs of TonB have been found in several Gram-negative species (14). The N terminus is in the cytoplasm; the protein is anchored in the inner membrane by its uncleaved N-terminal signal sequence (15, 16), and most of the protein extends into the periplasm. The membrane anchor sequence contains a set of highly conserved residues located on one face of the α-helix (SHLS motif). The sequence S*XXX*H (where *X* is any amino acid) has been defined as the minimal structural requirement for the coupling of TonB to the electrochemical gradient of the inner membrane (17). The amino acid sequence of TonB contains a long central region with a high percentage of proline residues between residues 70 and 102 (17%), which is thought to confer to TonB the conformational rigidity and extended shape necessary to span the periplasm, and thereby to allow the C-terminal domain to contact the receptor embedded in the OM (18). Mutational studies have defined the last 48 residues as being essential to make contact with the OM receptor (19).

TonB forms a complex in the inner membrane with ExbB and ExbD (13), two membrane proteins that could potentially act as proton translocators. ExbB is homologous to the protein MotA, and ExbD has a similar topology as MotB, both of which are thought to exploit the proton gradient to drive the bacterial flagellum. ExbB has been proposed to modulate the conformation of TonB (20), as well as mediate its recycling (12, 13), but it has remained an enigma as to what these structural changes

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Both authors contributed equally to this work.

‖ To whom correspondence should be addressed: Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Tel.: 41-1-635-5570; Fax: 41-1-635-5712; E-mail: plueckthun@biocfebs.unizh.ch.

[1] The abbreviations used are: OM, outer membrane; pmf, proton motive force; Mes, 4-morpholinoethanesulfonic acid; r.m.s., root mean square; IEX, ion exchange chromatography; IMAC, immobilized metal ion affinity chromatography; MAD, multiwavelength anomalous diffraction.

TABLE I
Data collection statistics

	Remote 2	Remote 1	Peak	Inflection
Wavelength (Å)	0.9800	0.9163	0.9196	0.9199
Resolution (Å)	20–1.55	20–2.0	20–2.0	20–2.0
Total reflections	138,987	80,451	81,133	58,622
Unique reflections	21,518	19,195	19,199	19,128
		(F+ and F− separated)	(F+ and F− separated)	(F+ and F− separated)
Completeness (%) (last shell)	96.8 (80.2)	99.6 (98.3)	99.6 (97.6)	99.1 (96.5)
R_{merge}	2.7 (11.0)	2.3 (4.1)	2.5 (4.3)	1.7 (4.2)

might be. Cross-linking studies have suggested the regions through which TonB might interact with its partners in the inner membrane: The contact with ExbB is mediated by the signal anchor (20), whereas the residues responsible for the interaction with ExbD are unknown (21).

TonB and its associated proteins ExbB·ExbD thus play the role of an energy-transducing complex, coupling the electrochemical proton gradient of the inner membrane to active import processes across the OM (13, 22). The energy is provided by the proton motive force (10, 23, 24). For the transduction process to occur, the C-terminal domain of TonB must contact the OM receptor. Based on genetic (25, 26), cross-linking (19, 27–29), and affinity chromatography (30) studies, a recognition site has been suggested on the receptor, the TonB box, a hydrophobic stretch of seven amino acids, which is highly conserved in all the TonB-dependent OM receptors (31). A recent study resulted in the proposal that the conformation rather than the sequence of the TonB box is important for the recognition process between TonB and the receptor (19). Moreover, it has been hypothesized that other regions of both interacting partners are also involved (27). Most strikingly, TonB dependence is maintained if the complete cork domain is deleted (32, 33), including the TonB box. It follows that the recognition site cannot be limited to the TonB box.

A number of phages and colicins have exploited the TonB·ExbB·ExbD system for gaining entry into bacteria (34). A similar system, TolQRA, has also been described as allowing entry for other phages and colicins (34). The cellular function of the TolQRA system has remained enigmatic, and its deletion leads to a leaky phenotype (although no such effect is caused by the deletion of TonB·ExbB·ExbD). Nevertheless, both systems can partially complement each other (35). We have recently described the crystal structure of the C-terminal domain of TolA (36), and we became interested in finding out whether any structural similarity might exist between the C-terminal domains of both TonB and TolA.

In this paper, we present the crystal structure of the C-terminal domain of TonB at 1.55-Å resolution and show that this protein exhibits a novel fold that is without homology to any known structures. Moreover, we provide the first evidence that the C-terminal domain of TonB forms a tightly intertwined dimer.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The sequence encoding residues 155–239 of *tonB* from *E. coli* strain JM83 was polymerase chain reaction-amplified and cloned into the plasmid pAT37 (based on pQE30 from Qiagen). pAT37 codes for protein D (gpD) from bacteriophage λ with an N-terminal His-tag, under control of the T5 promoter (37). The *tonB* gene was fused to the C terminus of gpD, with an enterokinase cleavage site engineered by the polymerase chain reaction in between the two proteins. Recombinant bovine enterokinase (purchased from Invitrogen) cleaves after the sequence Asp-Asp-Asp-Asp-Lys. The recombinant protein was expressed overnight at 30 °C in *E. coli* XL1-Blue, after induction with 1 mM isopropyl-β-D-thio-galactopyranoside. Cells were lysed with a French press and, after centrifugation, the gpD-TonB fusion protein remained in the soluble fraction. The undigested fusion was purified at pH 8.0, using the coupled IMAC-IEX (cation exchange) protocol (38) on a BIOCAD-60 workstation. After dialysis against 50 mM Tris, pH 8.0, 1 mM $CaCl_2$, 0.1% Tween 20, the cleavage reaction was performed at room temperature for 4 h, using 1 unit per mg of fusion of the recombinant bovine enterokinase (Invitrogen). The solution was then dialyzed against 50 mM Mes/Hepes/acetate buffer, pH 8.0. Removal of gpD and enterokinase was again achieved with the coupled IMAC-IEX (cation exchange) protocol, based on the different pI of TonB, gpD, and enterokinase.

Crystallization and Structure Solution—The C-terminal domain of TonB was dialyzed against 20 mM Tris buffer at pH 7.5 and was concentrated to 15 mg/ml. Crystallization was performed by the hanging-drop vapor diffusion method at 22 °C. Crystal screen I (Hampton Research) was used for the initial screening. Small, rod-shaped crystals were found under conditions 6, 19, 27, and 36. The final refined crystallization conditions were 28–30% polyethylene glycol 3350, 0.1 M Tris buffer at pH 7.5, 50–100 mM $CaCl_2$. After refinement of the conditions, crystals were grown to the size of 0.3–0.5 mm. When a crystal was picked up from a droplet, the diffraction pattern showed split spots or high mosaicity. To improve their quality, crystals were moved from the droplet to a well containing mother liquor and stored for more than 1 day. Such treatment both increased the resolution of diffraction and lowered the mosaicity. TonB crystals were found to belong to the orthorhombic space group $P2_12_12$ with the unit cell parameters a = 63.78 Å, b = 86.34 Å, c = 26.56 Å. The asymmetric unit contains two molecules, and the V_M value is 1.89 Å^3/Da (solvent content 35%).

The structure of TonB was solved by derivatization with Br^- ions (39, 40). To prepare a crystal for this procedure, it was soaked for 50 s in a solution containing 1.0 M KBr in addition to the crystallization buffer. Subsequently, the crystal was picked up with a nylon loop (Hampton Research) and was flash-frozen in a nitrogen stream. All data sets were collected at 100 K using the ADSC Quantum 4 charge-coupled device detector on the synchrotron beamline X9B at the National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY. The bromine fluorescence edge was scanned to determine the energy of the inflection, peak, and remote points. Three data sets were measured at 2.0-Å resolution to provide all information necessary for a multiwavelength anomalous diffraction (MAD) experiment. In addition, a data set extending to 1.55 Å was obtained for the purpose of structure refinement. Data were integrated and scaled using the HKL2000 program suite (41). Data collection statistics are summarized in Table I.

Four Br^- sites were found by both direct and Patterson methods and were refined using the program SOLVE (42), utilizing three data sets corresponding to the peak, inflection, and remote wavelengths, in the resolution range 10–2.5 Å. These sites were also confirmed with the program SHELXD (43). The phases from SOLVE were modified and extended to 1.55 Å using the program DM (44) in the CCP4 program suite (45), with the solvent content set at 25%. The mean figure of merit of the phase set was 0.608 for the 10–2.5 Å data after SOLVE, and 0.489 for 20–1.55 Å after DM (0.780 for 20–2.5 Å). The initial model was built using the automatic model-building option of the program ARP/wARP (46) with the full-DM phase set as input. The model was rebuilt with the program O (47) using either electron density maps based on the combination of the MAD and model phases, or straight $2F_o - F_c$ maps. The combined phase set was obtained using SIGMAA in the CCP4 program package. After each cycle of rebuilding, the model was refined using SHELXL (48) at the resolution range of 20 to 1.55 Å, without applying any non-crystallographic (NCS) restraints, as the latter prevented proper convergence. Eight full cycles of remodeling and refinement were performed, with the refinement of individual anisotropic B-factors for all atoms initiated in cycle six. In addition to protein atoms, 219 water molecules and four bromide ions have been added to the model. The R-value for all reflections between 20 and 1.55 Å is 16.0% (R_{free} 23.0%). The geometrical properties of the model were assessed by the

program PROCHECK (49), and the secondary structure elements were assigned by the program PROMOTIF (50). The figures were prepared with Molscript (51) or Bobscript (52) and rendered with Raster3D (53).

RESULTS

The crystal structure of the C-terminal domain of TonB has been determined by multiwavelength anomalous diffraction, and has been refined using SHELXL at 20–1.55 Å, yielding a model with low *R*-factor and excellent stereochemistry. The refinement statistics and the indicators of model quality are listed in Table II. The electron density maps (both the combined map utilizing the phases of the MAD data and of the model, and the final $2F_o - F_c$ map) are generally of excellent quality (Fig. 1). However, both maps are poorly defined in the neighborhood of residues 194–201. In this region, *B*-factors of all atoms are relatively high, indicating extensive flexibility of the polypeptide chain. Some disorder is also present at both termini of each molecule. Residues that are not visible in the maps include the first ten N-terminal amino acids of our construct (residues 155–164), as well as the last one or two residues on the C terminus (residues 238 and 239 of molecule A, and 239 of molecule B). The electron density of the remaining parts of the protein is very well defined. The mean positional error in atomic coordinates as estimated by the Luzzati plot is 0.16 Å. All non-glycine and non-proline residues of the model lie either in the most favorable region or in the additionally allowed region of the Ramachandran plot.

TABLE II
Refinement statistics for the final coordinates of the C-terminal domain of TonB

Resolution range	20.0–1.55 Å
Unique reflections used	20,365
R_{cryst}	16.0%
R_{free}	23.0% (5% test set)
r.m.s. deviations from ideality	
Bond lengths	0.009 Å
Angles	0.028 Å
Non-zero chiral volumes	0.054 Å^3
Zero chiral volumes	0.047 Å^3
Number of amino acid residues	73 + 74
Number of protein atoms	578 + 586
Number of heteroatoms	4
Number of solvent atoms	219

The C-terminal domain of TonB is cylinder-shaped with the length of 65 Å, and the diameter of 25 Å, with two protein chains forming a single compact entity. Each chain is rich in β-strands (~50% of the secondary structure) with much more limited extent (~15%) of residues found in helical conformation (either α-helix or 3_{10} helix). Each monomer contains three β-strands (strand S1, residues 169–182; S2, 188–194; and S3, 221–236) and one α-helix (residues 200–210 in molecule A, and 200–211 of molecule B). In addition, a short 3_{10} helix includes residues 211–213 of molecule A. All six β-strands make a large antiparallel β-sheet. The β-strand S3 of each monomer is

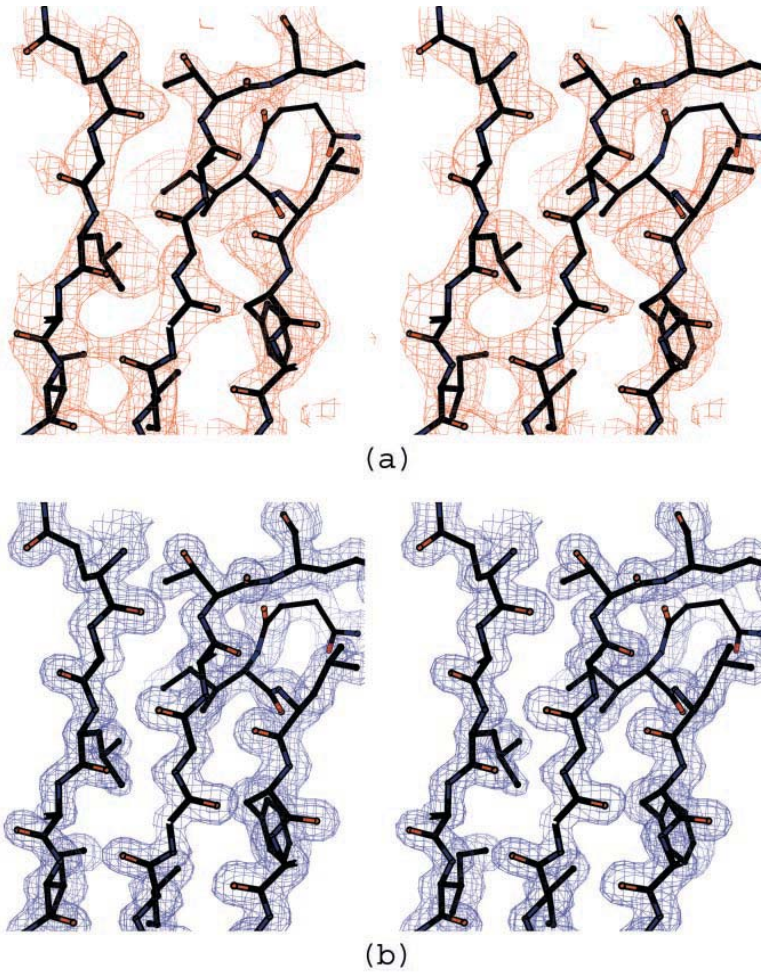


FIG. 1. **Electron density maps of the central β-sheet region of TonB.** *a*, stereoview of the MAD-phased electron density map contoured at 1.0 σ. The map was calculated with phases from the program SOLVE (42), modified with DM (44). *b*, stereoview of the final $2F_o - F_c$ map calculated with the program SHELX (48), contoured at 1.5 σ.

swapped between the monomers (Fig. 2). Four β-strands, S1 and S3 of both molecules, are located on one side, whereas the two short β-strands S2 and the helices are located on the other side (Fig. 3).

The two monomers differ slightly from each other. At any refinement stage, application of non-crystallographic restraints between the two molecules resulted in significantly worse behavior than if such restraints were not utilized. The r.m.s. deviation between the Cα atoms of the two monomers is 0.42 Å, whereas the r.m.s. deviation between the side-chain atoms of the two monomers is 1.15 Å. In the case of the main chain of the protein, the largest differences are found at the N terminus. As judged by their high temperature factors, both termini are located in highly flexible regions of the structure. The difference between Cα positions of Ala-165, the first visible residue on the N terminus, exceeds 2 Å (residues preceding Ala-165 were not visible in either molecule). In the case of the side chains, several residues show significantly different values of the χ_1 angle. These residues include Leu-170, Arg-171, Glu-173, Asn-200, Lys-219, and Lys-231. The only hydrophobic amino acid among them is Leu-170, and different orientation of its side chain leads to the presence of more hydrophobic contacts in molecule B. The other residues are all polar or charged, and all are solvent-exposed. The orientations of the Cγ atoms in the residues belonging to the β-sheet (Arg-171, Glu-173, Lys-231) closely coincide. For positively charged residues (Arg-171 and Lys-231 of both molecules) Cγ atoms point toward the N terminus of molecule A. The Cγ atoms of the negatively charged residue Glu-173 point toward the N terminus of molecule B. Asn-200 is located just after the highly flexible loop and is itself flexible, judged by its high *B*-factor. Lys-219 is located at the end of the molecule and is also flexible.

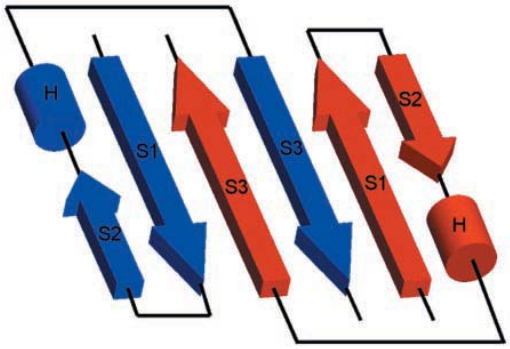


FIG. 2. **Schematic diagram of the secondary structure topology of the C-terminal domain of TonB.** Molecule A is colored in *red* and molecule B in *blue*. The *arrows* represent β strands, and each *cylinder* represents a helix. The secondary structure elements are labeled, and the residues belonging to each of them are described in the text.

The interactions between the two protein chains that form a single compact molecule of the C-terminal domain of TonB are unusually extensive. The dimeric interface area covers 41% of the surface of each monomer, thus the individual chains are unlikely to be able to exist independently and the protein becomes stable only as a dimer. The region of the β-sheet shows tight dimeric interactions, whereas the interactions on the opposite side of the molecule are not as close. Although the single antiparallel β-sheet present in the dimer is composed of strands originating from different molecules, the hydrogen bonding pattern is close to ideal. The loop between β-strand S2 and the α-helix is very flexible, as indicated by its high *B*-factor. The average *B*-factors of the main-chain atoms in this loop are 71 $Å^2$ and 60 $Å^2$ for molecules A and B, respectively, as compared with the respective averages for other areas of 20.6 $Å^2$ and 21.7 $Å^2$. Crystal packing in the vicinity of these loops is rather loose, resulting in the formation of clefts or channels on the molecular surface. The channels are made by residues 195–200 and 172–176 in both molecules, the former belonging to the loop, and the latter to the β-strand S1.

DISCUSSION

The binding of a nutrient, such as vitamin B_{12} or an iron·siderophore complex, to the external face of an outer membrane receptor triggers a series of conformational changes: The N-terminally located TonB box, which is hidden within the barrel of the unliganded receptor, is made to project in an extended form into the periplasm and, thus, becomes freely accessible for interaction with the C-terminal domain of TonB (28). Additionally, subtle structural changes of the receptor observed crystallographically, such as the upward translation of selected loops of the cork domain (also termed "hatch region"), may disrupt hydrophobic interactions between the so-

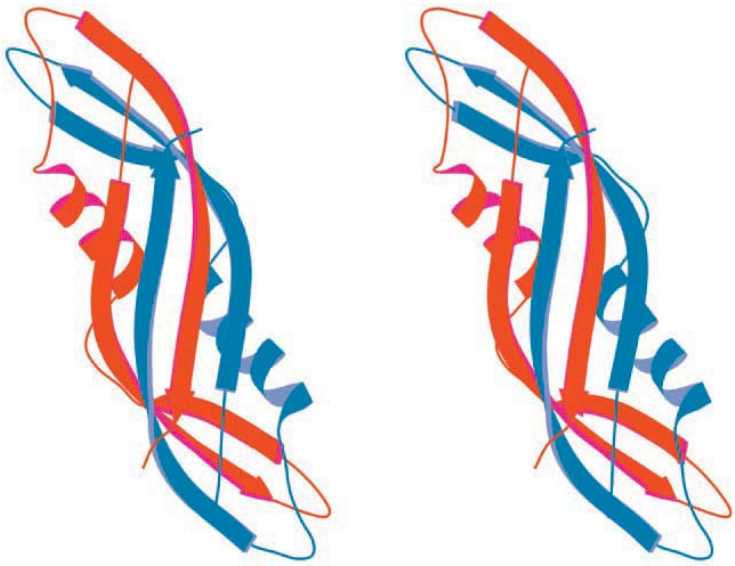


FIG. 3. **Stereo ribbon diagram of the C-terminal domain of TonB, showing the intertwined dimer.** The *color scheme* is the same as in Fig. 2. The atomic coordinates have been deposited in the Protein Data Bank (accession code 1IHR).

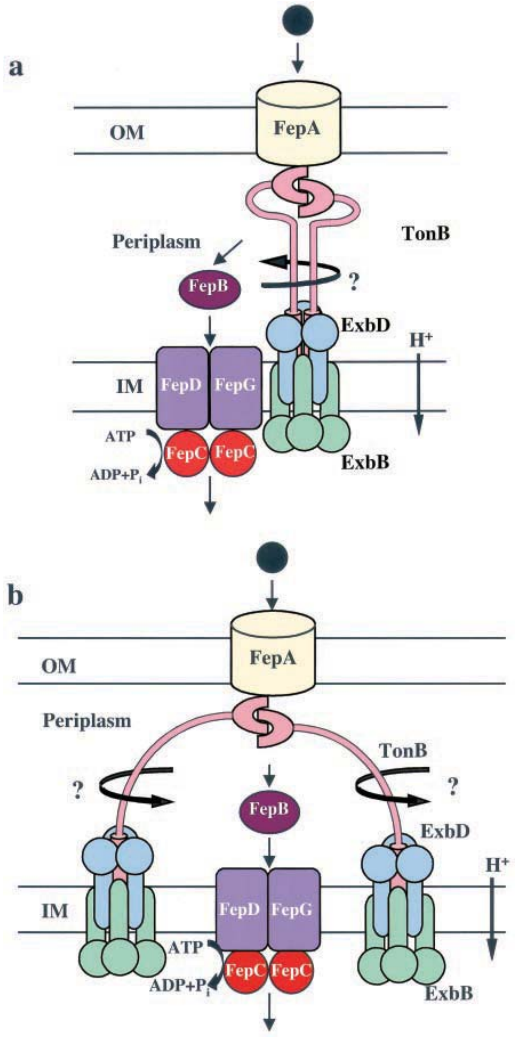


FIG. 4. **Hypothetical models of the interaction of dimeric TonB with (*a*) a single complex of trimeric ExbB·ExbD or (*b*) two complexes of ExbB·ExbD.** For details, see text.

called switch helix (residues 24–29 in FhuA) and the internal wall of the barrel, and the helix unfolds (6). It thus constitutes a candidate for signaling the occupancy of the outer binding site by a ligand to the periplasm. However, the actual "tag" being recognized by TonB could still be the TonB box, the switch helix mediating its accessibility. Nevertheless, there must be additional crucial conformational changes occurring in the barrel itself. Both in FhuA and FepA, the cork domain has been completely deleted and the TonB dependence of transport was fully maintained (32, 33). It follows that TonB must interact with the barrel itself, but this result certainly does not refute an interaction with the TonB box or other regions in the plug domains in wild-type receptors.

It is not clear whether substrate transport normally involves the complete dissociation of the plug domain from the barrel. On the one hand, this does not seem to be necessary, because it is conceivable that channels of sufficient dimension can be created by much smaller movements and changes in the cork domain, even though this is a matter of debate (5, 6). On the other hand, phages can inject their DNA through this pore (34) and colicins use it for entry (34), and this is only conceivable with a completely unplugged pore. Furthermore, the unplugged state is apparently functional (32, 33) and thus able to exist, and therefore a full opening of the pore is likely at least for the entry of very large molecules.

The ligand-mediated signal could therefore trigger a conformational rearrangement first at the loops of the cork domain and the barrel of the receptor which is then transmitted along the barrel. The binding of TonB, which appears to be not continuous but to occur in cycles (12), may then stabilize an intrinsically energetically unfavorable conformation of the barrel, which allows the passage of the ligand. Additionally, TonB may bind to the cork domain in wild-type receptor and help its dislocation, but this is apparently not the decisive action for mediating ligand transport. The binding of TonB to the barrel is needed to effect ligand translocation. The "energizing" of the receptor might then simply consist of the binding of TonB to an intrinsically unstable form of the barrel, which stabilizes this form, allowing the passage of the nutrient, and a subsequent release of the TonB-receptor interaction is needed. We do not know which of these steps would require energy, and it might conceivably be the dissociation of TonB. Although this is ultimately a mechanical act, more sophisticated possibilities exist for a polypeptide machine than simple rigid movement. The barrel domain would then return to its ground state, ready to accept the next ligand molecule. An "energized conformation" of the TonB C-terminal domain would not be required in our model. We consider the possibility (see below) that this transient binding of TonB to the receptor barrel involves a rotary motion in the cytoplasmic membrane.

Cross-linking studies have indicated that the region around residue 160 of TonB is crucial for the interaction with the TonB box (27, 29, 54). In our model, this region is not visible because of its very high flexibility (see the description of the *B*-factors in "Results"). However, it is very likely that other regions of TonB are involved in the contact with the receptor, and it is clear that there must be an interaction with the barrel. Based on NMR studies, it has also been proposed that the C-terminal portion of the proline-rich segment (Lys-Pro repeats) (amino acids 91–102) of TonB is involved in a specific interaction with FhuA (55), but the binding constant is very weak, and this region has been deleted, with TonB still maintaining activity (56). We postulate that the region around Asn-200 of TonB (N terminus of the α helix) constitutes a binding cleft, which could accommodate an element of the receptor as the ligand, based on the relatively high conformational lability of this segment (see "Results"). Interestingly, the α-helix-forming residues (Met-201 through Arg-214) are conserved among several Gram-negative species (14). However, it is unlikely that the role assigned to this otherwise correctly predicted amphipathic helix by Larsen *et al.* (56), namely, binding to the outer membrane, is correct, because the hydrophobic side of this helix faces the core of the dimer and thus cannot participate in any other interactions.

Once the substrate reaches the periplasmic side of the receptor (*e.g.* FepA), it is taken up by the periplasmic-binding protein FepB. The subsequent steps of import are not well known. Substrate-containing FepB might then bind to the ABC transporter $FepC_2DG$, resulting in transport across the inner membrane, using energy derived from ATP hydrolysis (3). FhuA uses an analogous system with FhuD as a periplasmic-binding protein and FhuBC anchored in the inner membrane (57, 58).

In the present work, we provide the first evidence that the C-terminal domain of TonB forms a dimer. However, this dimerization that involves almost half of the surface of this protein domain does not correlate with the recent model, showing a homotrimeric ExbB·ExbD complex interacting with a

TonB monomer, which in turn contacts the OM receptor (21). Furthermore, the antiparallel orientation of each monomer as well as the cylindrical shape do not correspond to any topological representation of TonB described so far in the literature. Recently, a soluble form of TonB was expressed, which lacks the N-terminal anchor helix but contains the full proline-rich region (59). The authors interpreted equilibrium sedimentation and gel filtration data as indicating mostly monomers, even though the measured molecular weight was somewhat higher than expected.

The tightly intertwined dimeric structure of TonB seen in the crystal now leads to two possible, albeit speculative models (Fig. 4). In one case (Fig. 4*a*), both TonB proteins interact with the same ExbB·ExbD complex. It is attractive to hypothesize that the proton gradient might cause a torsional motion, as is found in several molecular machines such as the flagellum or ATP synthases, because of some homology in ExbB to MotA and a similar topology of ExbD to MotB. Two proline-rich regions would provide a stiffer structure than only one, which could thus directly transduce this force to the TonB C-terminal domain and mediate its transient interaction with the receptor barrel and/or the cork domain. Alternatively (Fig. 4*b*), each TonB monomer might be linked to a different ExbB·ExbD complex, yet a torsion of both might still be mechanically transduced to the outer membrane. Postle and co-workers (56) also deleted most of the proline-rich region, yet TonB was still functional. Nevertheless, a torsional mechanism could still be operational in these short-necked constructs. In our model, TonB only needs to bind, and dissociate again, but in a cyclic manner—we cannot distinguish which is the energy-requiring step. Undoubtedly, further work is necessary to test the validity of either model, and particularly the arrangement in the inner membrane needs to be clarified.

One of the goals of this project was to determine if any structural homology exists between the C-terminal domains of TonB and TolA (36). A direct comparison of these domains proves that their structure is completely different. Moreover, a comparison of the structure of TonB with all known protein folds (60) detected no structural relationship between the C-terminal domain of TonB and any other structure represented in the Protein Data Bank. This was true regardless of whether a monomer or a dimer of TonB was utilized as a search model. In this respect, the structure described here represents a totally new fold that has never been observed so far.

In conclusion, the most surprising finding of our structure is that the C-terminal domain of TonB forms a rigid and tightly intertwined dimer. It is conceivable that this is essential for transducing a mechanical force from the inner to the outer membrane, and it would be much harder to imagine this to occur with a monomeric molecule. Even though polyproline stretches have extended conformations, they are still flexible and are typical hinge regions, as exemplified in IgG molecules. It would thus be difficult to visualize how mechanical energy can be transduced with a flexible tether of a monomeric molecule. Our dimeric structure provides now a framework for further probing of the mechanism of the TonB-dependent import.

Acknowledgments—We thank Dr. Z. Dauter for his assistance in data collection on beamline X9B at the National Synchrotron Light Source, Brookhaven National Laboratory, and Dr. P. Forrer for providing the gpD fusion system for use in the expression of TonB.

REFERENCES

1. Schirmer, T. (1998) *J. Struct. Biol.* **121,** 101–109
2. Ehrmann, M., Ehrle, R., Hofmann, E., Boos, W., and Schlosser, A. (1998) *Mol. Microbiol.* **29,** 685–694
3. Buchanan, S. K. (2001) *Trends Biochem. Sci.* **26,** 3–6
4. Buchanan, S. K., Smith, B. S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D., and Deisenhofer, J. (1999) *Nat. Struct. Biol.* **6,** 56–63
5. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (1998) *Science* **282,** 2215–2220
6. Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P., and Moras, D. (1998) *Cell* **95,** 771–778
7. Newton, S. M., Allen, J. S., Cao, Z., Qi, Z., Jiang, X., Sprencel, C., Igo, J. D., Foster, S. B., Payne, M. A., and Klebba, P. E. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4560–4565
8. Nikaido, H., and Saier, M. H., Jr. (1992) *Science* **258,** 936–942
9. Jiang, X., Payne, M. A., Cao, Z., Foster, S. B., Feix, J. B., Newton, S. M., and Klebba, P. E. (1997) *Science* **276,** 1261–1264
10. Reynolds, P. R., Mottur, G. P., and Bradbeer, C. (1980) *J. Biol. Chem.* **255,** 4313–4319
11. Wooldridge, K. G., Morrissey, J. A., and Williams, P. H. (1992) *J. Gen. Microbiol.* **138,** 597–603
12. Larsen, R. A., Thomas, M. G., and Postle, K. (1999) *Mol. Microbiol.* **31,** 1809–1824
13. Letain, T. E., and Postle, K. (1997) *Mol. Microbiol.* **24,** 271–283
14. Larsen, R. A., Myers, P. S., Skare, J. T., Seachord, C. L., Darveau, R. P., and Postle, K. (1996) *J. Bacteriol.* **178,** 1363–1373
15. Postle, K., and Skare, J. T. (1988) *J. Biol. Chem.* **263,** 11000–11007
16. Karlsson, M., Hannavy, K., and Higgins, C. F. (1993) *Mol. Microbiol.* **8,** 379–388
17. Larsen, R. A., and Postle, K. (2001) *J. Biol. Chem.* **276,** 8111–8117
18. Skare, J. T., Ahmer, B. M., Seachord, C. L., Darveau, R. P., and Postle, K. (1993) *J. Biol. Chem.* **268,** 16302–16308
19. Larsen, R. A., Foster-Hartnett, D., McIntosh, M. A., and Postle, K. (1997) *J. Bacteriol.* **179,** 3213–3221
20. Larsen, R. A., Thomas, M. G., Wood, G. E., and Postle, K. (1994) *Mol. Microbiol.* **13,** 627–640
21. Higgs, P. I., Myers, P. S., and Postle, K. (1998) *J. Bacteriol.* **180,** 6031–6038
22. Braun, V. (1995) *FEMS Microbiol. Rev.* **16,** 295–307
23. Hancock, R. W., and Braun, V. (1976) *J. Bacteriol.* **125,** 409–415
24. Bradbeer, C. (1993) *J. Bacteriol.* **175,** 3146–3150
25. Schramm, E., Mende, J., Braun, V., and Kamp, R. M. (1987) *J. Bacteriol.* **169,** 3350–3357
26. Gudmundsdottir, A., Bell, P. E., Lundrigan, M. D., Bradbeer, C., and Kadner, R. J. (1989) *J. Bacteriol.* **171,** 6526–6533
27. Cadieux, N., and Kadner, R. J. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10673–10678
28. Merianos, H. J., Cadieux, N., Lin, C. H., Kadner, R. J., and Cafiso, D. S. (2000) *Nat. Struct. Biol.* **7,** 205–209
29. Cadieux, N., Bradbeer, C., and Kadner, R. J. (2000) *J. Bacteriol.* **182,** 5954–5961
30. Moeck, G. S., Coulton, J. W., and Postle, K. (1997) *J. Biol. Chem.* **272,** 28391–28397
31. Postle, K. (1993) *J. Bioenerg. Biomembr.* **25,** 591–601
32. Braun, M., Killmann, H., and Braun, V. (1999) *Mol. Microbiol.* **33,** 1037–1049
33. Scott, D. C., Cao, Z., Qi, Z., Bauler, M., Igo, J. D., Newton, S. M., and Klebba, P. E. (2001) *J. Biol. Chem.* **276,** 13025–13033
34. Lazdunski, C. J., Bouveret, E., Rigal, A., Journet, L., Lloubes, R., and Benedetti, H. (1998) *J. Bacteriol.* **180,** 4993–5002
35. Eick-Helmerich, K., and Braun, V. (1989) *J. Bacteriol.* **171,** 5117–5126
36. Lubkowski, J., Hennecke, F., Plückthun, A., and Wlodawer, A. (1999) *Structure* **7,** 711–722
37. Forrer, P., and Jaussi, R. (1998) *Gene* **224,** 45–52
38. Plückthun, A., Krebber, A., Krebber, C., Horn, U., Knüpfer, U., Wenderoth, R., Nieba, L., Proba, K., and Riesenberg, D. (1996) in *Antibody Engineering: A Practical Approach* (McCafferty, J., and Hoogenboom, H. R., eds) pp. 203–252, IRL Press, Oxford
39. Dauter, Z., Dauter, M., and Rajashankar, K. R. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 232–237
40. Dauter, Z., Li, M., and Wlodawer, A. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 239–249
41. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
42. Terwilliger, T. C., and Berendzen, J. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52,** 749–757
43. Sheldrick, G. M. (1997) *Methods Enzymol.* **276,** 628–641
44. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 487–493
45. Collaborative Computational Project Number 4. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
46. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
47. Jones, T. A., and Kieldgaard, M. (1997) *Methods Enzymol.* **277,** 173–208
48. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol.* **277,** 319–343
49. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
50. Hutchinson, E. G., and Thornton, J. M. (1996) *Protein Sci.* **5,** 212–220
51. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
52. Esnouf, R. M. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 938–940
53. Meritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 869–873
54. Kadner, R. J. (1990) *Mol. Microbiol.* **4,** 2027–2033
55. Brewer, S., Tolley, M., Trayer, I. P., Barr, G. C., Dorman, C. J., Hannavy, K., Higgins, C. F., Evans, J. S., Levine, B. A., and Wormald, M. R. (1990) *J. Mol. Biol.* **216,** 883–895
56. Larsen, R. A., Wood, G. E., and Postle, K. (1993) *Mol. Microbiol* **10,** 943–953
57. Braun, V., Hantke, K., and Köster, W. (1998) in *Metal Ions in Biological Systems: Iron Transport and Storage in Microorganisms, Plants and Animals* (Sigel, A., and Sigel, H., eds) pp. 67–145, Marcel Dekker, New York
58. Groeger, W., and Koster, W. (1998) *Microbiology* **144,** 2759–2769
59. Moeck, G. S., and Letellier, L. (2001) *J. Bacteriol.* **183,** 2755–2764
60. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138

2.3 New results on TonB and the ExbB-ExbD pathway

2.3.1 Discussion

The question of the oligomeric state of TonB *in vivo* and *in vitro* has been a subject of intense discussions since the publication of our dimeric structure. The scepticism displayed in the beginning by certain authors has been rapidly cleared up, and based on data reported by several groups it has become increasingly clear that the C-terminal domain (residues 155-239) forms a dimer *in vitro*, and that this is not due to a crystallization artifact. This is now widely accepted among researchers of the field, and we can reasonably say that our structure is correct and reflects the true *in vitro* state of TonB in solution. Molecular dynamics simulations have even been undertaken to confirm our finding (see http://indigo1.biop.ox.ac.uk/josed/thesis/chapter6.pdf, unpublished results). However, many authors found that our model did not fit with their data. As first pointed out by Letellier et al., a longer fragment (residues 33-239) clearly behaved as a monomer in sedimentation analyses (Moeck and Letellier, 2001). By analyzing TonB proteins of several lengths, Koedding et al. (2004) came to the same conclusion: a short C-terminal fragment of TonB encompassing residues 164 to 239 (TonB-77) or 155 to 239 (TonB-86) formed homodimers in solution, whereas longer fragments (starting at residues 145, 135 or 125) remained monomeric (Koedding et al., 2004). Of all the fragments investigated by these authors, it was found that TonB-77 was not able to bind to FhuA in vivo, whereas all the others constructs did. They conclude that the dimeric structures of TonB-77 and TonB-86 are not the *in vivo* functional state, because the longer fragments were monomeric and inhibited TonB-dependent transport far more effectively. Sauter at al. (2003) came to the same conclusion: although TonB (164-239) formed a dimer in vivo when fused to ToxR (cholera toxin gene), they did not find any evidence of dimer formation with a fusion between a longer TonB (residue 33-239) and ToxR, thereby proposing that a region between residues 33 and 164 prevents dimer formation (Sauter et al., 2003). But they further note that if the full TonB (1-239) is fused to ToxR, then the ability to form dimer is restored, thus showing the importance of the transmembrane region in dimer formation of full-length TonB. It appears then that TonB contains both N-proximal and C-proximal dimerization modules. From all the above mentioned studies, one might at first draw the conclusion that our TonB molecule is simply not functional.

But the reality is perhaps a bit more complex, especially when the interaction with the outer membrane receptor (e.g. FhuA) is considered (Figure 2-15). A proposed mechanism of FhuA-TonB interaction postulates that when the dimeric C-terminal TonB (residue 155-239) contacts FhuA, an “encounter complex” is formed with a certain affinity (~840 nM), which is lowered to form a functional complex upon addition of the ligand (ferricrocin). However, in the case of the full periplasmic fragment (residue 32-239), the “encounter complex” might involve two TonB monomers bound to FhuA at both the N-terminal and C-terminal domains of each TonB monomers, thus yielding a bound dimer of TonB. Upon addition of ferricrocin, the affinity of the N-terminal domain of TonB for FhuA increases; the affinity of the C-terminal domain decreases (Khursigara et al., 2004). Similar experiments were performed in Susan Buchanan’s group in order to co-crystallize TonB with

the purified outer membrane receptor FecA. A his-tagged FecA (plus or minus ferric citrate) was first incubated with “our” TonB, and the mixture was applied onto a Ni-NTA column, the idea being to capture TonB and bind the whole complex to the column via the FecA his-tag. However, a mixture of FecA and TonB was present in the flowthrough, and only uncomplexed FecA bound to the column. Similarly, it was not clear whether the co-crystals may contain TonB. The same set of experiments has been repeated with Cir (colicin I receptor), again without success (Susan Buchanan, personal communications). All these considerations let us to suppose that “our” TonB does not interact with either transporter strongly enough to isolate a complex, which led some authors to suggest that as a result of crystallization in the absence of the amino-terminus, our structure might represent an intermediate unenergized form of TonB (Ghosh and Postle, 2004). In agreement with this suggestion, other authors (Sean Peacock et al., 2005) have solved by NMR the structure of a longer TonB fragment (residue 103-239) and found it to be monomeric, including several structural aspects that are different from the previously published crystal structures. They postulate that both the monomer and dimer forms of TonB may contribute to the energy transduction process across the periplasm.

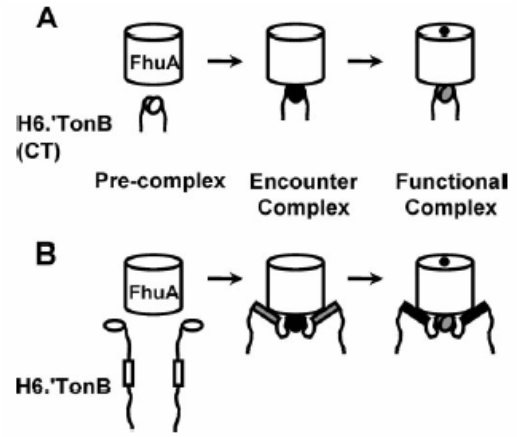


Figure 2-15. Proposed mechanism of FhuA-TonB interactions. In panel A, H6.’TonB (CT) is the hexahistidine-tagged C-terminal TonB, residues 155-239. Panel B involves a hexa-histidine-tagged full length TonB consisting of residues 32-239. The oval shapes represent the C-terminal domain for both H6.’TonB (CT) and H6.’TonB, whereas the rectangular shape represents the N-terminal domain found in H6.’TonB only. Differences in shading represent the modulations in affinity caused by the natural ligand (ferricrocin). Dark gray shading indicates a stronger affinity than light gray. Picture taken from Khursigara et al. (2004).

Our understanding of the processes governing transduction of the proton motive force (pmf) energy from the inner to the outer membrane still remains far from complete. Two models have emerged in the last few years, and they are discussed in greater details in

in a recent microreview (Postle and Kadner, 2003). The propeller model (Figure 4, our publication) states that TonB remains permanently associated with the inner membrane, undergoing a rotary motion of the carboxy-terminal propeller, that is initiated at the inner membrane by ExbB, ExbD and the pmf. In the shuttle model, uncharged TonB in complex at the inner membrane with ExbB/D is converted to charged TonB by the passage of a proton through the ExbB/D complex. Charged TonB shuttles to the outer membrane where it docks with the outer membrane receptor complexed with its ligand. After C-terminal contact with the outer membrane, the TonB N-terminus is released from the ExbB/D complex and would associate with the outer membrane transporter. Ligand binding results in a conformational change in the transporter, which induces a productive interaction with charged TonB and release of conformationally stored potential energy from TonB. Available data suggest that, although the inner membrane pmf is required to energize TonB, it does not energize the shuttling between membranes. A possible energy source for this “back-and-forth” movement between membranes could be ATP hydrolysis. The shuttle model has been tested in vivo (Larsen et al., 2003) and the data support the interpretation that, at some point during the transduction cycle, TonB is wholly associated with the outer membrane, having relinquished its inner membrane anchor. The single objection these authors formulate to their model is that the requirement for the hydrophobic transmembrane domain of TonB to leave and reinsert into the inner lipid bilayer is a thermodynamically unfavourable process.

Since the publication of our paper, the ferric dicitrate transporter FecA (Ferguson et al., 2002) and the cobalamin transporter BtuB (Chimento et al., 2003) have been crystallized, allowing a broader description of the structural features that are common or specific to each transporter. The basic modular architecture of the cobalamin transporter BtuB is similar to that of the iron siderophore transporters FepA, FhuA and FecA; specifically, an N-terminal hatch domain is embedded within the lumen of a 22-stranded β-barrel. The strands are tilted by around 45° relative to the membrane bilayer. The barrels are generally elliptical in cross-section, but are not superimposable. The FecA structure reveals that the ligand, in this case ferric citrate, binds to transporter and closes the central channel behind it, preventing access to the external milieu. This finding generated the air-lock model, suggesting that the loops occlude the channel to prevent passage of external material that might come across the outer membrane into the periplasm along with the transported substrate. Comparison of the structures of the three loaded transporters (FhuA, FecA and BtuB) shows that the relative locations of the binding sites for three very dissimilar substrates are nonetheless very similar in their location. BtuB is between 60-100 residues shorter than the iron siderophore transporters solved so far. BtuB has a more squat structure, possessing shorter extracellular loops and containing a smaller hatch domain than these other transporters. From the structure, it is clear that the loops in BtuB are unable to intertwine and block the channel from the outside, as seen for FecA. The structures of the hatches are highly conserved among all transporters, and the interface between hatch and barrel is mainly through polar interactions, with no large hydrophobic patches. Investigations of the physical properties of the hatch domain alone, as it may exist if it is removed from the FepA barrel, established that it behaves *in vitro* as a predominantly unfolded yet soluble protein (Usher et al., 2001). The finding that the hatch domain alone retained some specificity for FepA ligand (ferric enterobactin), though 100-fold lower than for the wild-type FepA receptor (5 μM against 50 nM for intact FepA), is consistent with its retaining biological

function outside of the barrel. However, the authors conclude that it remains entirely possible that the hatch domain does not come out of the barrel at all, but instead undergoes a major conformational rearrangement within the barrel to form a channel large enough to allow the ligand to pass by. They further propose that larger objects such as colicins cross the membrane by some other mechanism.

The crystal structures for FhuA (Ferguson et al., 1998; Locher et al., 1998), FecA (Ferguson et al., 2002) and BtuB (Chimento et al., 2003) bound and unbound to their cognate ligand have been solved. In all three cases, binding of ligand to the external face of the transporter causes a large conformational shift in the region of the protein at the periplasmic face. Intriguingly, this conformational change does not alter the location of the globular hatch domain or create a passageway through which ligand can exit. It has long remained a challenging question whether transport involves the complete or partial dissociation of the hatch domain from the barrel. Two models have been proposed (Figure 2-16), each having its major drawback (Usher et al., 2001).

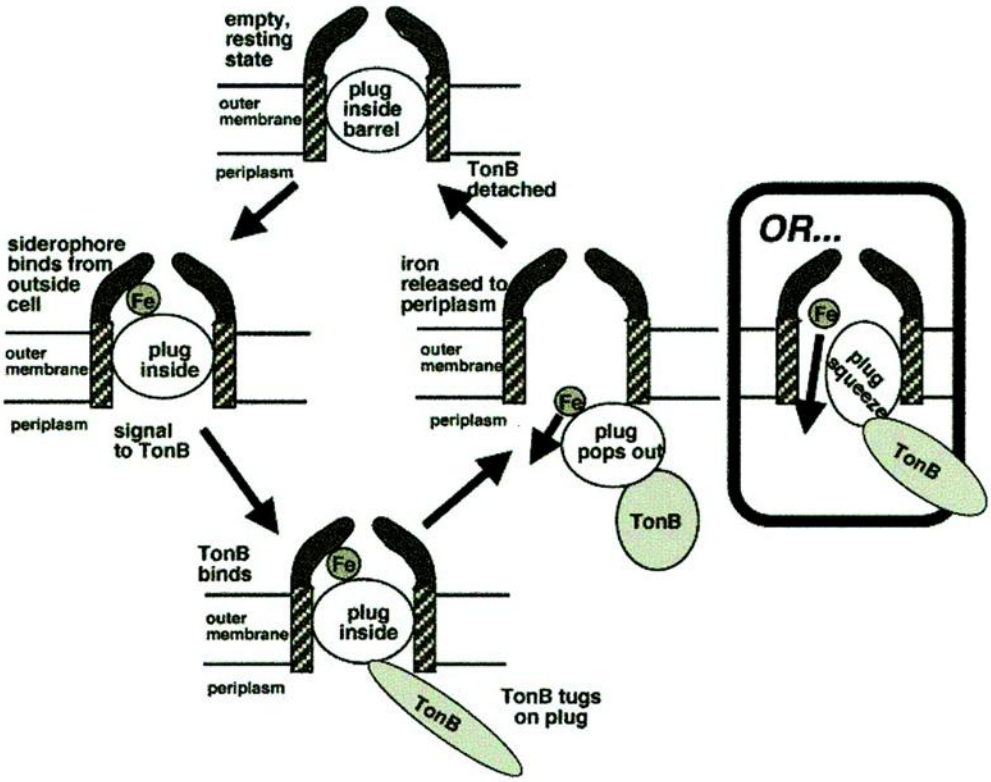


Figure 2-16. Schematic diagram illustrating critical steps in the transport cycle of siderophores by FepA (or other TonB-dependent receptors). The FepA transporter resides in the outer membrane of E. coli; the plug domain folds up inside the barrel and completely occludes the barrel's interior (upper cartoon). Upon ligand binding (e.g. siderophore), a signal is transmitted to TonB. This model proposes that TonB pulls the unfolded plug domain out of the barrel entirely or that TonB causes a major conformational rearrangement of the plug domain within the barrel, thereby forming a channel large enough to allow the passage of the ligand. Picture taken from Usher et al. (2001).

The model presented in Figure 2-16 is consistent for small molecules such as enterobactin or siderophores. However, it does not explain how larger objects such as colicins can cross the outer membrane. An attractive idea was recently presented in the view of the crystal structure of colicin E3 R-domain bound to BtuB. The authors provide evidence that we might have been entirely confused; these outer membrane transporters, according to their views, do not form pores and are not gated open by their ligands, which suggests that they should be called TonB-gated transporters to reflect the way they work. This statement was used to propose a new transport mechanism, implying the OmpF porin as a translocator (Figure 2-17).

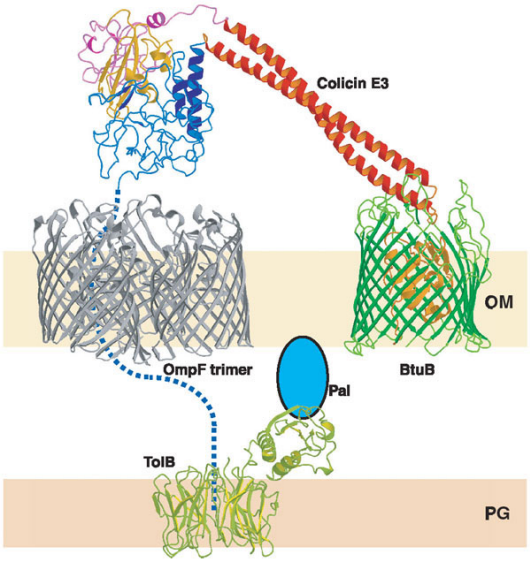


Figure 2-17. Model for an interaction between BtuB and OmpF that would be mediated by the colicin E3 bound to BtuB. The different colors in the colicin E3 symbolize the translocation (blue), catalytic (magenta) and immunity (yellow) domains. OmpF is a trimer in the outer membrane. Picture taken from Kurisu et al. (2003).

The Ton box is a short, moderately conserved stretch of residues near the amino-terminus of most TonB-dependant transporters. It is a key factor in signaling and transport, and is so far the only site of mutations that confer a TonB-uncoupled phenotype. Several lines of evidence point to interaction of the Ton box with TonB, which does not exclude that other sites in TonB or in the transporter may be in contact. This region is the site of mutations, such as the substitution of Pro for Leu-8 or Val-10 in BtuB, which result in loss of cobalamin transport but do not affect cobalamin binding or entry of the TonB-independent phage BF23 or E colicins. The Ton box could be modeled in the recent crystal structure of BtuB, but only minor changes were observed after cobalamin binding (Chimento et al., 2003). Using site-directed spin labeling of the N-terminal region of BtuB, the Kadner lab presented some direct evidence for the substrate-induced structural transition of the Ton box (Cadieux et al., 2003; Fanucci et al., 2003a). Their data support a consistent scenario in

which the Ton box conformation exists in equilibrium between a docked and undocked state, and that the binding of the translocation domain of colicin E3 shifts this equilibrium toward the docked state. In this conformation, the Ton box exhibits low backbone mobility and low accessibility to labeling agents. This stabilization ensures that there is no interaction between BtuB and TonB, and that there is no opening of the transport pore. This finding suggests that the colicin probably does not cross the outer membrane through the BtuB barrel, consistent with the model depicted in Figure 2-17. Another BtuB substrate, cyano-cobalamin, had the opposite effect on accessibility of the Ton box; its binding results in undocking, which is characterized by high residue mobility and high labeling. The increasing disorder of positions 6 to 8 of the Ton box upon cyano-cobalamin binding has also been observed in the crystal structure (relatively high B-factors of the Cα in the structure solved with the ligand). It is somewhat striking that both these substrates have opposite effects on the energetics of the Ton box segment of BtuB. An interesting structural challenge will be to determine how two opposite conformations of the Ton box are propagated through the BtuB barrel to the periplasm. Additional data about the structure and mobility of the Ton box were reported, indicating that, in its docked state, the Ton box is predominantly in a β–strand or extended conformation. Loss on conformational constraint upon cyano-cobalamin binding affects residues 1 to 15 of the BtuB core, and residues 16 and 17 appear to be near a hinge responsible for opening the Ton box (Fanucci et al., 2003b).

In recent years, important advances have furthered our understanding of the sophisticated and highly efficient process of siderophore uptake in gram-negative bacteria. Among these, it is worth noting the use of spectroscopic techniques (e.g. EPR, SDSL) to monitor the uptake process *in vivo*, the progression in the characterization of the Ton complex and, most notably, the determination of the atomic structure of several outer-membrane receptors. However, it seems that fundamental questions concerning many aspects of the transport process remain to be resolved. It is still not clear, what parts of TonB are involved in interaction with the transporters: the extreme C-terminus is important for function, but the sequences show low homology. The conformational changes that the outer-membrane receptors must undergo to allow permeation are also unknown, and how the hatch domain reorients to allow passage of substrate is still a matter of controversy. Similarly, the mechanism by which these changes are energetically activated from the inner membrane is poorly understood, as is the dependance of the Ton complex on the inner-membrane proton gradient.

2.3.2 Discussion

Cadieux, N., Phan, P. G., Cafiso, D. S., and Kadner, R. J. (2003). Differential substrate-induced signaling through the TonB-dependent transporter BtuB. Proc Natl Acad Sci USA 100, 10688-10693.

Chimento, D. P., Mohanty, A. K., Kadner, R. J., and Wiener, M. C. (2003). Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. Nat Struct Biol 10, 394-401.

Fanucci, G. E., Cadieux, N., Kadner, R. J., and Cafiso, D. S. (2003a). Competing ligands stabilize alternate conformations of the energy coupling motif of a TonB-dependent outer membrane transporter. Proc Natl Acad Sci USA 100, 11382-11387.

Fanucci, G. E., Coggshall, K. A., Cadieux, N., Kim, M., Kadner, R. J., and Cafiso, D. S. (2003b). Substrate-induced conformational changes of the periplasmic N-terminus of an outer-membrane transporter by site-directed spin labeling. Biochemistry 42, 1391-1400.

Ferguson, A. D., Chakraborty, R., Smith, B. S., Esser, L., van der Helm, D., and Deisenhofer, J. (2002). Structural basis of gating by the outer membrane transporter FecA. Science 295, 1715-1719.

Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (1998). Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science 282, 2215-2220.

Ghosh, J., and Postle, K. (2004). Evidence for dynamic clustering of carboxy-terminal aromatic amino acids in TonB-dependent energy transduction. Mol Microbiol 51, 203-213.

Khursigara, C. M., De Crescenzo, G., Pawelek, P. D., and Coulton, J. W. (2004). Enhanced binding of TonB to a ligand-loaded outer membrane receptor: role of the oligomeric state of TonB in formation of a functional FhuA.TonB complex. J Biol Chem 279, 7405-7412.

Koedding, J., Howard, P., Kaufmann, L., Polzer, P., Lustig, A., and Welte, W. (2004). Dimerization of TonB Is Not Essential for Its Binding to the Outer Membrane Siderophore Receptor FhuA of Escherichia coli. J Biol Chem 279, 9978-9986.

Kurisu, G., Zakharov, S. D., Zhalnina, M. V., Bano, S., Eroukova, V. Y., Rokitskaya, T. I., Antonenko, Y. N., Wiener, M. C., and Cramer, W. A. (2003). The structure of BtuB with bound colicin E3 R-domain implies a translocon. Nat Struct Biol 10, 948-954.

Larsen, R. A., Letain, T. E., and Postle, K. (2003). In vivo evidence of TonB shuttling between the cytoplasmic and outer membrane in Escherichia coli. Mol Microbiol 49, 211-218.

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P., and Moras, D. (1998). Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. Cell 95, 771-778.

Moeck, G. S., and Letellier, L. (2001). Characterization of in vitro interactions between a truncated TonB protein from Escherichia coli and the outer membrane receptors FhuA and FepA. J Bacteriol 183, 2755-2764.

Postle, K., and Kadner, R. J. (2003). Touch and go: tying TonB to transport. Mol Microbiol 49, 869-882.

Sauter, A., Howard, S. P., and Braun, V. (2003). In vivo evidence for TonB dimerization. J Bacteriol 185, 5747-5754.

Sean Peacock, R., Weljie, A. M., Peter Howard, S., Price, F. D., and Vogel, H. J. (2005). The Solution Structure of the C-terminal Domain of TonB and Interaction Studies with TonB Box Peptides. J Mol Biol 345, 1185-1197.

Usher, K. C., Ozkan, E., Gardner, K. H., and Deisenhofer, J. (2001). The plug domain of FepA, a TonB-dependent transport protein from Escherichia coli, binds its siderophore in the absence of the transmembrane barrel domain. Proc Natl Acad Sci USA 98, 10676-10681.

SECTION 3
Application of phage display

Introduction

The aim of this third section is to illustrate by a concrete example the usefulness of the current knowledge in filamentous phage biology. The growing amount of knowledge collected over the last decade has led some researchers to take profit of this wealth of data to design a new technology which has become an indispensable tool for protein engineering. Fifteen years have elapsed since the first demonstration that antibodies could be displayed on the surface of phage (McCafferty et al., 1990). During these years, the power of combining the technologies of phage display and protein chemistry has become increasingly apparent. In phage display, large combinatorial libraries of encoded proteins (in our case, recombinant antibodies) are displayed as fusions to the minor coat protein of the phage capsid. Phage displaying a protein with a desired property can be retrieved from the library and decoded by sequencing the encapsulated phage DNA. By this way, phenotype (the displayed protein) and genotype (DNA) are physically linked.

Nearly all of the work on protein engineering using phage display has been done using the filamentous phage M13 or the closely related phage fd. This is in part due to the ease with which this phage can be manipulated and also to our detailed understanding of the viral life cycle and phage structure (see section 1). In our set-up, the display of proteins on the phage surface is accomplished by inserting the DNA encoding the displayed protein (i.e. the genotype) between the signal sequence of the coat protein and the amino terminus of the C-terminal part of gene 3 protein. The signal sequence is proteolytically removed upon secretion into the periplasm of *E. coli* where phage assembly occurs.

The principle of selection we have applied is based on searching the antibody library for those members with the highest affinity for a particular target. To achieve this affinity maturation process, an off-rate selection has been applied. Phage displaying antibodies are allowed to bind biotinylated antigen in solution. By setting the concentration of the biotinylated antigen lower than the equilibrium dissociation constant (Hawkins et al., 1992), phage could be selected based on their relative binding strength. This method can be adapted to select for prolonged off-rate (k_{off}). After this initial binding event, a large excess of free unlabelled antigen is added resulting in dissociation of the phage from the biotinylated antigen (Jermutus et al., 2001; Zahnd et al., 2004). The rate of this reaction is proportional to k_{off} and the exposed antibody heads are captured by unlabelled antigen. The longer this

off-reaction is allowed to proceed before capture, the greater the selection for prolonged off-rate. The complexes of phage still bound to the biotinylated antigen are then captured on streptavidin-coated paramagnetic beads. After washing the beads, the phage are eluted and can be used for a next selection round.

Off-rate selection has been applied to a recombinant antibody directed against the murine MHC-gp33 peptide complex, in order to improve its moderate affinity constant (low micromolar range). The contrasting results of this study are presented and discussed in the following pages.

Biochemical characterization of a single-domain antibody selected against a murine MHC-peptide complex

3.1 Introduction

The immune system is a remarkably adaptive defense system that has evolved in vertebrates to protect them from invading pathogenic microorganisms. It is able to generate an enormous variety of cells and molecules capable of specifically recognizing and eliminating an apparently limitless variety of foreign invaders. The immune system is able to recognize subtle chemical differences that distinguish one foreign pathogen from another. Once a foreign organism is recognized, the immune system enlists the participation of a variety of cells and molecules to mount an appropriate response, known as an effector function, to eliminate or neutralize the organism. Viruses are such foreign invaders, able to divert the host's cellular machinery to translate their own proteins and replicate the viral genome. The strategy for recognition of these intracellular parasites by the cellular immune system involves the cell surface presentation of short segments of intracellularly processed proteins of the invading organism by major histocompatibility complex (MHC) class I molecules on the surface of the infected cell; this conversion of foreign proteins into MHC-associated peptide fragments is called antigen processing, and the composition of the pool on the surface of the infected cell mirrors its intracellular protein content (Yewdell, 2001; Yewdell and Bennink, 2001). These peptides usually derive from defective ribosomal translation products (Princiotta et al., 2003), but in a virally infected cell, MHC-presented peptides derive from viral proteins. More than 10'000 different peptides, of about 8 or 9 amino acids in length, are displayed on each cell. Infections by pathogens result in the appearance of foreign or "abnormal" peptides, which mark the cell for destruction by cytotoxic T lymphocytes (CTL). It has been known for some time that most of MHC class I-presented peptides are generated by proteolytic degradation of endogenous proteins. The proteasome, a cytoplasmic multi-unit structure has been identified as a key element in the processing pathway, responsible for the generation of the COOH terminus of the peptide (Niedermann, 2002). The protease trimming the NH_2 terminus of the peptide remained largely elusive, until its recent identification (Saric et al., 2002; York et al., 2002).

The MHC class I molecule consists of three distinct regions: cytoplasmic, transmembrane and extracellular. The latter is a heterodimer of a heavy chain, a 45'000 MW type I integral membrane glycoprotein, and β2-microglobulin (β2m), a 12'000 MW soluble protein. The extracellular region of the heavy chain folds into three domains (α1,α2, α3) with a noncovalently associated molecule of β2-microglobulin (Figure 3-10). X-ray crystallographic studies have revealed that the peptide-binding site on the MHC consists of a flat undulating groove, located on the upper surface of the molecule, between the two antiparallel α helices of the α_1/α_2 domain (Young et al., 1994).

T-cell receptors (TCRs) are heterodimers, composed of two protein chains ($\alpha\beta$) linked by disulfide bridges. Each chain consists of an N-terminal variable (V) region and a C-terminal constant (C) region. The Vα and Vβ regions are structurally related to the V domains of the heavy (H) and light (L) chains of immunoglobulins (Igs) and their mode of interaction is identical to that found in antibody VL-VH dimers. T-cell receptors scan the composite surfaces produced by the MHC molecules and “sense” the presence of peptides residing in their groove; this information is then relayed to the interior of T-cells triggering their proliferation (Valitutti et al., 1995). All TCRs analyzed to date bind to the peptide-MHC complex (pMHC) with roughly similar diagonal geometry (Hennecke and Wiley, 2001; Housset and Malissen, 2003). A major difference in antigen recognition by antibody as compared with TCR is that the former commonly functions through high-affinity interactions, whereas the strength of the TCR interaction with pMHC is weaker (Willcox et al., 1999).

Detailed knowledge of the formation of the pMHC complex and the nature of its binding to the TCR are prerequisites for a complete understanding of the cellular and molecular processes governing antigen processing and presentation on the surface of the infected cell. However, to date very few tools have been developed to directly detect, visualize, count and study peptidic antigen present on an infected cell. To fulfill this role, there have been attempts to use soluble and specific TCRs engineered from cloned T cells of known pMHC specificity (Holler et al., 2003; Plaksin et al., 1997). Although engineered TCRs seem to be ideal for this purpose, it has been proven practically difficult to engineer these molecules because their inherent low affinity for their target ligands and their instability as recombinant-engineered molecules may limit their use as detection reagents (Wülfing and Plückthun, 1994). Clearly, there is a growing need for engineering new targeting molecules, able to recognize with high affinity and specificity pMHC complexes. Antibodies represent the most suitable strategy for such an approach, and they have already been used to study antigen presentation (Day et al., 1997; Porgador et al., 1997; Zhong et al., 1997), to localize and quantify antigen-presenting cell displaying a T cell epitope (Inaba et al., 1997), specifically mask an autoimmune T cell epitope (Aharoni et al., 1991; Puri et al., 1997) or as a targeting tool in a mouse model (Reiter et al., 1997). Such unique antibodies with T-cell receptor-like specificity may prove to be a valuable tool not only for obtaining precise information about the presence, expression pattern and distribution of the pMHC complex, but they may also serve as good candidates for targeting reagents in constructing recombinant immunotoxins or for antibody-based therapeutic approaches. Using very large nonimmune phage antibody libraries, the direct selection of such antibodies with exquisite specificity was only recently demonstrated (Chames et al., 2000; Denkberg et al., 2002). Detailed in Figure 3-10, the crystal structure of an affinity-matured Fab bound to an HLA-A1:MAGE-A1 complex supports the evidence that antibodies and TCR share a characteristic diagonal docking mode for most peptide-MHC interactions (Hülsmeyer et al., 2005). Another group has taken advantage of this technique to select high affinity human Fab antibodies against a tumor-associated antigen (Lev et al., 2002).

The scFv_A8 was originally isolated from HuCAL, a fully synthetic antibody library displayed on filamentous phage (Knappik et al., 2000). It was selected against the extracellular part of a murine MHC-gp33 peptide complex. Four panning rounds of phage display against the MHC-gp33 complex permitted the rapid identification of a few candidate

clones. One of them (scFv_A8) showed promising results with regard to its specificity and expression behaviour (Frédéric Pécorari, unpublished results), and was converted to another antibody format (Fab fragment, see example structure in Figure 3-10), thus resulting in Fab_A8. As shown by surface plasmon resonance studies (BIACORE), Fab_A8 has an affinity of 1 μM for the MHC-gp33 complex (Maurice Brozzo, diploma thesis, Zürich 2000). Based on previous observations (Ewert et al., 2003a), six amino acid mutations were introduced into the variable domain of the heavy chain (VH6 framework), thus markedly improving the thermodynamical properties and the expression behaviour of this domain. VH6 domains (as well as VH2 and VH4) exhibited low cooperativity during denaturant-induced unfolding, lower expression yields, and higher tendency to aggregate (Ewert et al., 2003a). In all subsequent chapters, I will refer to this particular clone Fab_A8 as the **ConFab**, meaning the non-improved Fab fragment containing the wild-type amino acid sequence.

In the present work, we describe the *in vitro* affinity maturation of the ConFab, conducted using phage display methods. Genetic diversity is introduced into the gene encoding ConFab by error-prone PCR; variants showing apparently improved affinity for the MHC-gp33 complex were isolated, and we provide evidences that, using a particular antibody format, the main features of the ConFab could be maintained. The isolation of a single-domain antibody is detailed and its slightly improved binding properties are illustrated in comparison with the parental ConFab.

3.2 Materials & Methods

Construction of phage libraries and mutagenesis

We made use of the plasmid pMorph18 (Morphosys, Martinsried, Germany), a phagemid construct with chloramphenicol resistance containing the gene encoding the ConFab, as the parental ("wild-type") construct for libraries described here. In subsequent chapters, we will refer to this clone as the "control Fab" (or **ConFab**). The libraries are fused as an N-terminal fusion to the C-terminus of gene 3 protein (pIII) from bacteriophage M13 (Figure 3-1a). Randomized libraries were generated by error-prone PCR (epPCR) amplification of the ConFab template, allowing control over the number of mutations per gene (Zaccolo et al., 1996). The whole variable domain of the light chain (Vκ2 subtype) was amplified with the primers Vκ2_for (5'-TGCACTGCAGAAGCAGCCAAA-3') and Vκ2_rev (5'-AAGGCCTTCGTACGTTTAATTTCAACTT-3'). The whole variable domain of the heavy chain (VH6 subtype) was amplified with the following primers:

5'-TAACGCTACAATTGGTACAGTCTGGT-3' (VH6_for)

5'-TAATGGCCTTGGCCCCAATAATC-3' (VH6_rev)

Standard PCR conditions were used for the amplification, except that 6-(2-deoxy-β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimidino-[4,5-c][1,2]oxazin-7-on-triphosphate (dPTP) and 8-oxo-2'-deoxyguanosine (8-oxo-dGTP) were added (Jena Biosciences, cat.# NU-1119S and NU-1117S, respectively). Two libraries were designed, containing 5 (low error) and 15 (high error) mutations per clone on average. Based on 20 PCR cycles the nucleotide analogues dPTP and 8-oxo-dGTP were added at 1/10 (40 μM total) and 3/10 (120 μM total) of the total concentration of dNTPs (400 μM) for the low and high error library, respectively. The PCR products were purified by gel extraction, digested using restriction enzymes PstI/Bsi WI (for Vκ2 library) and MfeI/StyI (for VH6 library). The restricted and purified PCR products were ligated into the pMorph18 phagemid, digested with the corresponding restriction enzymes, electroporated into competent XL1-Blue F+ *E. coli* cells and amplified by growing 15-16 hours at 30°C. The library was rescued by overnight growth at 37°C, using the VCSM13 helper phage (Stratagene, Cat.# 200251) with a titer of 10^{10}. Phage particles were purified and concentrated by two PEG/NaCl precipitations (20% polyethyleneglycol 6000, 2.5M NaCl, sterile filtered), resuspended in 1ml PBS buffer (50 mM potassium phosphate, 150 mM NaCl, pH 7.5), and filtered through a 0.45μm sterile filter. The complexity of the libraries were estimated based on plating of the initial electroporation onto 2 x TY plates containing chloramphenicol (34 μg/ml)+1% glucose, and reached the following values: 10^5 (Vκ2), 10^6 (VH6 "low error"), $3x10^6$ (VH6 "high error"). The efficiency of error-prone PCR mutagenesis was checked by sequencing using the following primers (Figure 3-1): "M13 rev" for Vκ2 (5'-CAGGAAACAGCTATGAC-3') and "HuCAL for" for VH6 (5'-TACCGTTGCTCTTCACCCC-3'). Sequencing results are presented in Figure 3-2.

Off-rate selection

Each of the three libraries (Vκ2 "low error", VH6 "low error" and "high error") was treated separately for 2 rounds of phage display selections. All selections and washing steps were done at room temperature. Antigen solutions and magnetic beads are equilibrated in PBS buffer pH 7.4 before use. For each round, approximately 10^{11} phages (≈1 nM) displaying the library were screened for binding to biotinylated MHC-gp33, whose concentration was 100 nM and 50 nM for selection rounds 1 and 2 respectively. The mixture of phage and antigen, containing a 100x excess of β2-microglobulin was gently rotated for one hour. A large molar excess of non-biotinylated over biotinylated MHC-gp33 (50 times in round 1, 100 times in round 2) was then added in order to select for Fab fragments with the lowest dissociation constant (k_{off}). Competition was allowed to occur for 1 and 10 minutes (in round 1), 10 and 100 minutes (in round 2), before capturing the bound complexes with 200 μl of milk-blocked streptavidin magnetic beads (Roche, cat.# 1'641'778) in round 1, and 100 μl of milk-blocked avidin agarose beads (Sigma, cat.# A9207) in round 2. Capture with beads was done for 30 minutes, before washing them 9 times for 2 minutes: three times in PBS/2% milk, three times in PBS/0.1% Tween-20, three times in PBS. For elution, the beads were supplemented with 200 μl of 100 mM triethylamine and agitated for 8 minutes. After removal of the beads, the supernatant was neutralized with 200 μl of 1M Tris, pH7.4, and used to infect 10 ml log phage *E. coli* XLI Blue F+ cells (OD_{600} = 0.5). Subcloning of DNA from the final selection rounds into the plasmid pMorphx9_FH (from Morphosys, see Figure 3-1b) allowed soluble periplasmic expression under the control of the *lac* promoter in *E. coli* strain SB536.

ELISA with peptide-MHC complex

All antigens were coated directly onto Maxisorp plates (NUNC, cat.# 442404) in TBS_{150} (20 mM Tris, 150 mM NaCl, pH 8.0), overnight at 4°C at a concentration of 120 nM. Plates were washed three times with $TBST_{150}$ (TBS_{150} containing 0.05% Tween-20) and blocked with 5% milk-$TBST_{150}$ for 2 hours at room temperature. After 3 times washing, antibodies (as crude extract or purified protein) were added and incubated for 90 minutes at room temperature. Following 3 washes, detection was done either with mouse anti-Flag M2 (Sigma, cat.# F3165, 1:1'000 dilution) or mouse anti Tetra-His (Qiagen, cat.# 34670, at 1:1'000) in 2% milk-$TBST_{150}$. As a secondary antibody, a goat anti-mouse IgG conjugated to alkaline phosphatase was used (Sigma, cat.# A3562, at 1:5'000 in 2% milk-$TBST_{150}$). Color development was done with 4-NPP (3 mM 4-NPP, 50 mM $NaHCO_3$, 50 mM $MgCl_2$), and absorbance was measured at 405 nm with an HTS 7000 PLUS plate reader (Perkin Elmer).

Crude extracts of Fab were prepared as follows: His- and Flag-tagged Fab antibody fragments inserted into plasmid pMorphx7_FH were expressed overnight in SB536 strain of *E. coli* at 30°C under constant agitation, in a sterile Maxisorp plate (Nunc, cat.# 163320). Lysis was done by adding into each well 40 μl of BEL buffer (400 mM boric acid, 320 mM NaCl, 4 mM EDTA, pH8.0) containing 2.5 mg/ml lysozyme, and shaking for 1 hour at room temperature. Twenty microliters of this crude extract preparation were pipetted into coated wells of an ELISA plate.

For the phage ELISA, an anti-M13 antibody coupled to horseradish peroxidase was used (Amersham Pharmacia Biotech, cat.# 27-9421-01) and detection was done with soluble BM blue POD substrate (Roche Diagnostics, cat.# 1484281). Absorbance was measured at 405 nm with an HTS 7000 PLUS plate reader (Perkin Elmer). Phage particles were precipitated once with an ice cold PEG/NaCl solution (20% polyethyleneglycol 6000, 2.5 M NaCl, sterile filtered), redissolved in 1 ml PBS buffer and titrated before adding them into the ELISA wells.

Western blot of Fab fragments

15% SDS-PAGE gels were blotted onto Protran™ nitrocellulose membranes (Schleicher&Schuell, cat.# 10401196) for 1 hour under constant voltage (1 Volt/cm^2 membrane). All the following incubations were done in 5% $MTBST_{150}$ for 1 hour at room temperature: after blocking, the membrane was incubated with either an anti-human kappa light chain coupled to alkaline phosphatase (Southern Biotech, cat.# 2060-04, dilution 1:3'000) or an anti-TetraHis (1:1'000) antibody. A goat anti-mouse IgG coupled to alkaline phosphatase (1:5'000) was used as secondary antibody with anti-Tetra His. Detection was performed with 100 μl NBT (30 mg/ml in 70% DMF) and 100 μl BCIP (15mg/ml in 100% DMF) dissolved in 10 ml AP buffer (100 mM Tris pH 9.5, 0.5 mM $MgCl_2$). Color development takes 10 minutes.

Expression and purification of peptide-MHC complexes

Plasmids encoding H-$2D^b$ or β_2-microglobulin were transformed into *E. coli* strain BL21-CodonPlus-(DE3)-RIL and expression was carried out for 5 hours at 37°C. One liter of 2xTY medium (containing 100 μg/ml Ampicillin, 30 μg/ml Chloramphenicol, 0.1% glucose) was inoculated with 25 ml (for H-2Db) or 12 ml (for β2-m) of a saturated overnight culture. At on OD_{600} of 0.7-0.8, expression was induced by addition of a 1 mM IPTG solution (final concentration) and the cultures were incubated for 4 hours at 37°C, whereby they reached a typical OD_{600} of 4.5. The bacteria were pelleted (30 min, 20'000g, 4°C), resuspended in lysis buffer (50 mM Tris pH 8.0, 1 mM EDTA, 500 mM NaCl, 10 μg/ml DNAse I) and disrupted by French Press (Emulsiflex, Avestin) lysis. Inclusion bodies were pelleted by centrifugation (20 min, 39'000g, 4°C). The supernatant was discarded, the inclusion bodies were carefully resuspended by sonication (3 x 2 minutes) in 15 ml TE buffer (50 mM Tris, 1 mM EDTA, pH 8.0) containing 1% Triton X-100, and then centrifuged for 20 minutes (39'000g, 4°C). The supernatant was discarded and this washing procedure was repeated two times more until the supernatant was not yellow anymore. The washing was repeated one time with TE buffer alone, and two times with TE buffer containing 1 M guanidinium-HCl. The pellet of inclusion bodies was homogenously beige after this extensive washing procedure and solubilization could be started by adding the 10 ml solubilization buffer (50 mM Tris pH 8.0, 6 M guanidinium-HCl, 2 mM DTT (usually 10 ml is enough for a pellet resulting from 1 L culture). The inclusion bodies were resuspended by sonication (2 x 3 minutes) and placed at room temperature with stirring overnight to complete solubilization. Remaining insoluble particles were pelleted by centrifugation (20 minutes, 39'000g). The denatured and reduced

H-2D^b subunit (≈32 mg) and β_2-m subunit (≈ 12mg) were refolded by rapid dilution in 1 liter of refolding buffer consisting of 100 mM Tris, 1 M L-arginine, 2 mM EDTA, 1 mM oxidized glutathione, 0.2 mM reduced glutathione containing 10 mg of gp33 or ASIA peptide (see sequences below). The refolding buffer was cooled at 4°C and the pH adjusted to 8.0. Refolding was performed at 4°C for 24 hours under stirring. The protein solution was concentrated to 50 ml with an Amicon 8400 concentrator and then dialyzed once against 30 mM MHA buffer pH 8.0 (10 mM MES, 10 mM HEPES, 10 mM sodium acetate). After elimination of insoluble material by centrifugation (20 minutes, 5'000g) and filtration (0.22 μm pore size), the peptide-MHC complex was purified over a strong anion exchange chromatography column (equilibrated with MHA buffer, pH 8.0) on a BioCAD perfusion chromatography workstation (Applied Biosystems). The bound protein was eluted using an NaCl gradient ranging from 150 mM to 2.4 M NaCl. Purity was checked on a 15% SDS-PAGE; the protein solution was dialyzed twice against TBS_{150} (20 mM Tris, 150 mM NaCl, pH 8.0), aliquoted and stored at -80°C for further use.

Peptides

Peptides were synthesized by Jerini (Jerini AG, Berlin, Germany) to 95% purity and checked for molecular mass by mass spectrometry analysis. The synthetic peptides had the following sequences: SALQNAASIA (ASIA peptide) and KAVYNFATC (gp33 peptide).

Biotinylation of the MHC-gp33 complex

The MHC-gp33 complex in TBS150, pH 8.0 was mixed in a 2 ml volume with 5 mM ATP, 10 μM biotin, 2.5 μM BirA, 50 mM Tris pH 7.4, 5 mM $MgCl_2$, and the reaction was allowed to proceed overnight at room temperature. End concentration of the MHC complex in the reaction was around 3 μM. The solution was then dialyzed twice against 1 L PBS, pH 7.4, and purified through a monomeric avidin column (Pierce). Free biotin was removed by dialysis against PBS buffer. The yield of the reaction after purification was around 20%. Biotinylation of the antigen was checked on an immunoblot using an anti-streptavidin antibody coupled to alkaline phosphatase (Roche, cat.# 1089161).
BirA was produced as a C-terminal fusion to protein D as follows: the gene coding for BirA was cloned between the restriction sites *KpnI/HindIII* of the vector pAT194, thus giving rise to the construct pAT212 (Forrer and Jaussi, 1998). After transformation into competent XLIBlue cells, the culture was grown at 37°C in LB medium (supplemented with 100 μg/ml ampicillin, 1% glucose) until an OD_{600} ≈ 1.0 was reached. Fifteen minutes before induction, 50 μM D-biotin (Sigma) was added to the medium. Expression was then induced with 0.4 mM IPTG during 4 hours, at 37°C (typical OD_{600} ≈ 3.0). Cells were harvested, resuspended in TBS_{500} (20 mM Tris pH 8.5, 500 mM NaCl, 50 μM D-biotin), disrupted by French Press and the soluble fraction loaded onto an IMAC column saturated with Ni^{2+}-ions (immobilized metal ion affinity chromatography) on a BIOCAD perfusion chromatography workstation (Applied Biosystems). Protein could be eluted from the column with 300 mM imidazol (pH 8.0), and the eluate was reloaded onto an anion exchange column. The flowthrough was collected and dialyzed against BirA buffer (50mM Tris pH7.5, 250mM KCl, 0.1 mM DTT),

supplemented with 5% glycerol after dialysis. The enzyme was stored at 4°C. The final yield was 10 mg per liter culture.

Cloning and expression of the single-domain VH6 antibody

The variable domain of the heavy chain of the ConFab was PCR amplified using the following primers:

5'-GGCCTTCCATGGGACAGGTGCAATTGGTACAGTC-3' (named sdVH6_for)

5'-CTATCAAGCTTATCAGTGATGGTGATGGTGATGAGATCCTGAGCTAACCGTCACCAGG-3'

(named sdVH6_rev), containing an His-tag (underlined).

The conditions for the PCR reactions were as follows: 92°C (2 min); then 30 cycles at 92°C (30 sec.); 58°C (50 sec.); 72°C (40 sec.); and a final elongation step of 10 min at 72°C. Product size: 372 nucleotides.

The PCR product was purified and digested overnight at 37°C using the restriction enzymes *NcoI/HindIII*. After purification, the restricted PCR product was inserted into the similarly restricted pTFT74 vector. The ligated product was transformed into chemically competent *E. coli* XLIBlue cells. The sequence was verified by nucleotide sequencing.

Expression was done in 2xTY medium in BL21(DE3) cells as follows: from freshly transformed colonies on an agar plate, a single clone was picked and used to inoculate a culture containing 1% glucose and 100 μg/ml ampicillin. After overnight growth at 37°C, the culture was diluted 100x into fresh LB medium (supplemented with 100 μg/ml ampicillin, but without glucose) and allowed to grow until an OD_{600} of 0.8 was reached. Protein expression was induced with 1 mM IPTG, at 37°C. After 6h growth, cells usually reached an OD_{600} of 4.8. Cells were then harvested, resuspended in TE buffer pH 8.0 (20 mM Tris, 1 mM EDTA) containing 10 μg/ml DNAse. Cell lysis was performed with a French Press (Emulsiflex, Avestin). After centrifugation (30 minutes at 20'000 rpm), the pellet was kept. The inclusion bodies were washed as described for the MHC (see above). Refolding was performed in the same way as done for the MHC, except that the refolding buffer contained a five times excess of reduced glutathione over oxidized glutathione (5 mM to 1 mM). After concentration, the refolding solution was dialyzed against 20 mM NaH_2PO_4 pH7.5, 500 mM NaCl, and purified using a Nickel-NTA affinity column (IMAC) on a BioCAD perfusion chromatography workstation (Applied Biosystems). The bound protein was eluted with 200 mM imidazol (pH 8.0), and the buffer was changed to TBS150 (pH 7.5) for immediate use. Three milligrams pure refolded protein could be obtained from 1 L culture. Concentrations were measured by UV absorption using an extinction coefficient of 49'560 M^{-1} cm^{-1}

3.3 Results

In the absence of any information regarding the specific amino acids which interact between ConFab and the MHC-gp33 complex, we decided to randomize the whole variable domain of the light and the heavy chain of ConFab, using error-prone PCR (see *Materials and Methods*). In analogy to previous examples in the literature (Boder et al., 2000; Zahnd et al., 2004), we reasoned that randomizing regions of an antibody outside its CDR loops might significantly increase the probability of finding a better binder. This was later exemplified by the three-dimensional crystal structure of the affinity-matured Fab "hyb3" directed against a human MHC variant: all of the 22 mutations in the light chain were located at the periphery of the variable domain, far from the binding interface (Hülsmeyer et al., 2005).

Three libraries were constructed; the variable domain of the light chain (Vκ2 subtype) has only been subjected to a mild randomization, thus yielding a "Vκ2 low error" library, whereas both a "VH6 high error" and a "VH6 low error" library of the variable domain of the heavy chain (VH6 subtype) were created. For each library, eight clones were randomly picked and sequenced. The average number of amino acid mutations per clone was 5 for the "Vκ2 low error" library, 7.5 for the "VH6 low error" library and 16 for the "VH6 high error" library. Mutations were equally spread between the CDR and framework regions, with the notable exception that the framework region before the CDR1 of the Vκ2 domain was left untouched by the randomization procedure (Figure 3-2).

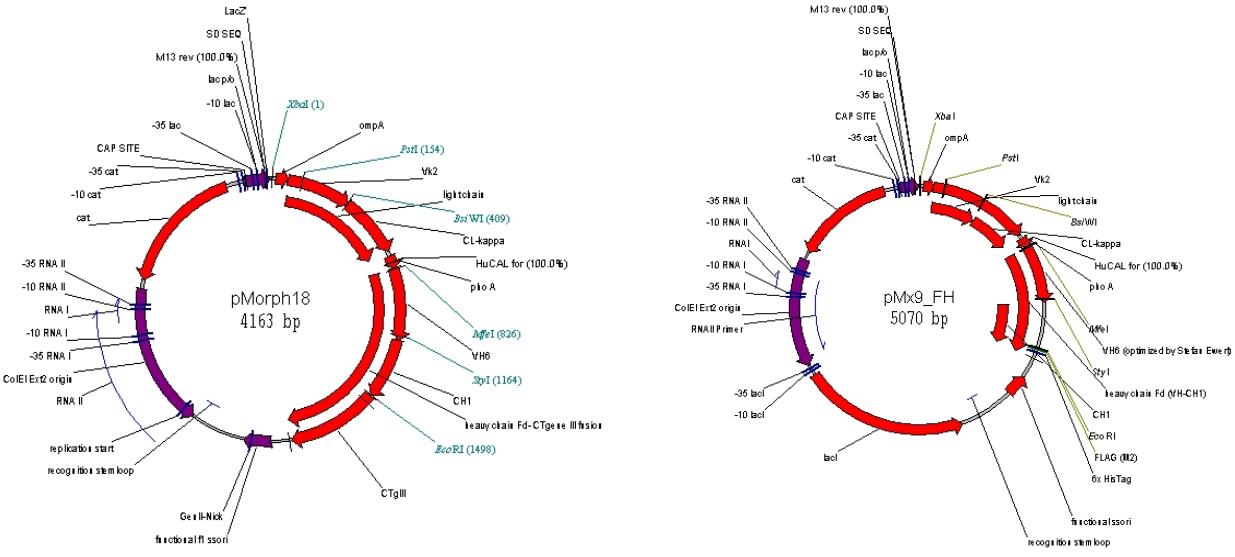


Figure 3-1. The pMorph18 display vector ("phagemid") is depicted on the left; the Fab library is fused to the amino terminus of the C-terminal domain of the minor coat protein (pIII) of phage M13. The restriction sites used for library cloning are indicated: *PstI/BsiWI* for Vκ2 and *MunI/StyI* for VH6. "cat" denotes the chloramphenicol resistance gene. pMx9_FH expression vector (right); the antibody is cloned in the same orientation, and antibiotic resistance is provided by the same antibiotic as in the pMorph18 system. M2 Flag- and His-tags are present at the C-terminus of the antibody fragment. Both vectors contain the *XbaI/EcoRI* restriction sites, allowing subcloning of the entire Fab fragment from one vector into another.

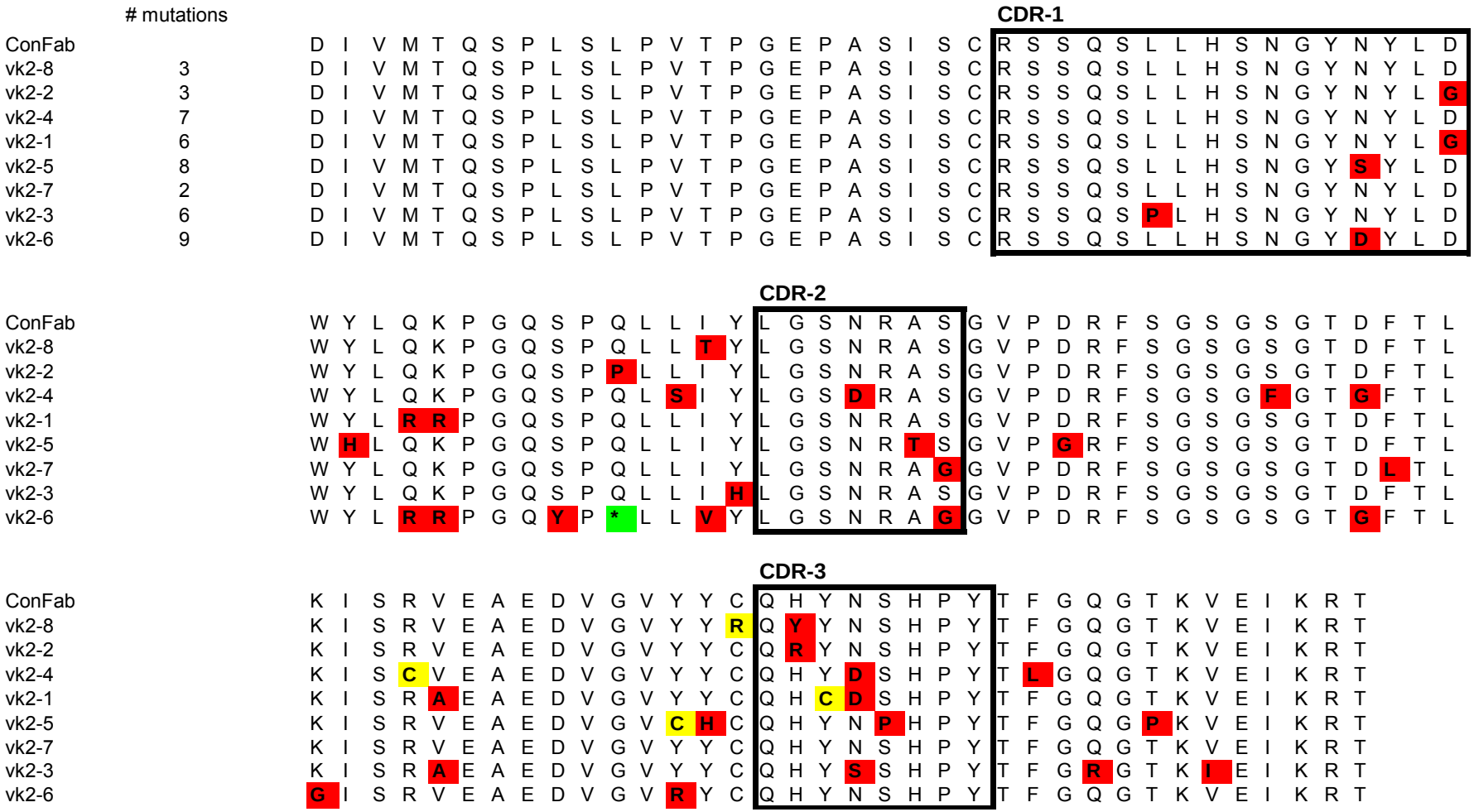


Figure 3-2. Alignment of unselected library members from the “Vk2 low error” library. Each randomly picked clone is denominated by a number (1 through 8). The second column indicates the number of mutation for each clone. The last clone (vk2-6) contains a STOP codon highlighted in green. Additional or lost cysteine residues are highlighted in yellow.

All three libraries were independently panned against the biotinylated MHC-gp33 complex bound to streptavidin magnetic beads (round 1) and avidin agarose beads (round 2). We first performed 2 rounds of selection, using decreasing biotinylated antigen concentration to favor selection of improved binders (Figure 3-3). Concomitantly, the concentration of the free competitor (non-biotinylated MHC-gp33) and the dissociation time were increased, as suggested elsewhere (Hawkins et al., 1992). After none of these first two rounds could any significant enrichment be observed: all three libraries showed enrichment factors not higher than 2- or 3-fold over a control selection containing only phage particles, magnetic beads and non-biotinylated MHC-gp33. However, a dramatic difference was observed in the second round of the Vκ2 selection; the number of phages eluted from the sample where off-rate selection was allowed for 10 minutes was nine times larger than the sample in which the off-rate was done for 100 minutes (Figure 3-3). Following these two rounds of initial off-rate selection, the DNA encoding the pool of selected Fabs with the randomized VH6 domains (from “high error” and “low error” libraries) was pooled, digested using *MunI*/*StyI* restriction enzymes and inserted into the phagemid encoding the selected Fabs with the randomized Vκ2 domain, digested with the same restriction enzymes. This procedure gave rise to a new library, termed “input Round 3” (see below). By this way, the randomized and selected variable domains of all three libraries (two from VH, one from VL) were genetically combined into the phagemid vector pMorph18. One additional selection round was performed with the “input round 3” library, but the clones isolated from this third selection round (output R3) could not be sequenced. We assume that some internal deletion occurred, thereby removing the annealing sites of the sequencing primers (data not shown).

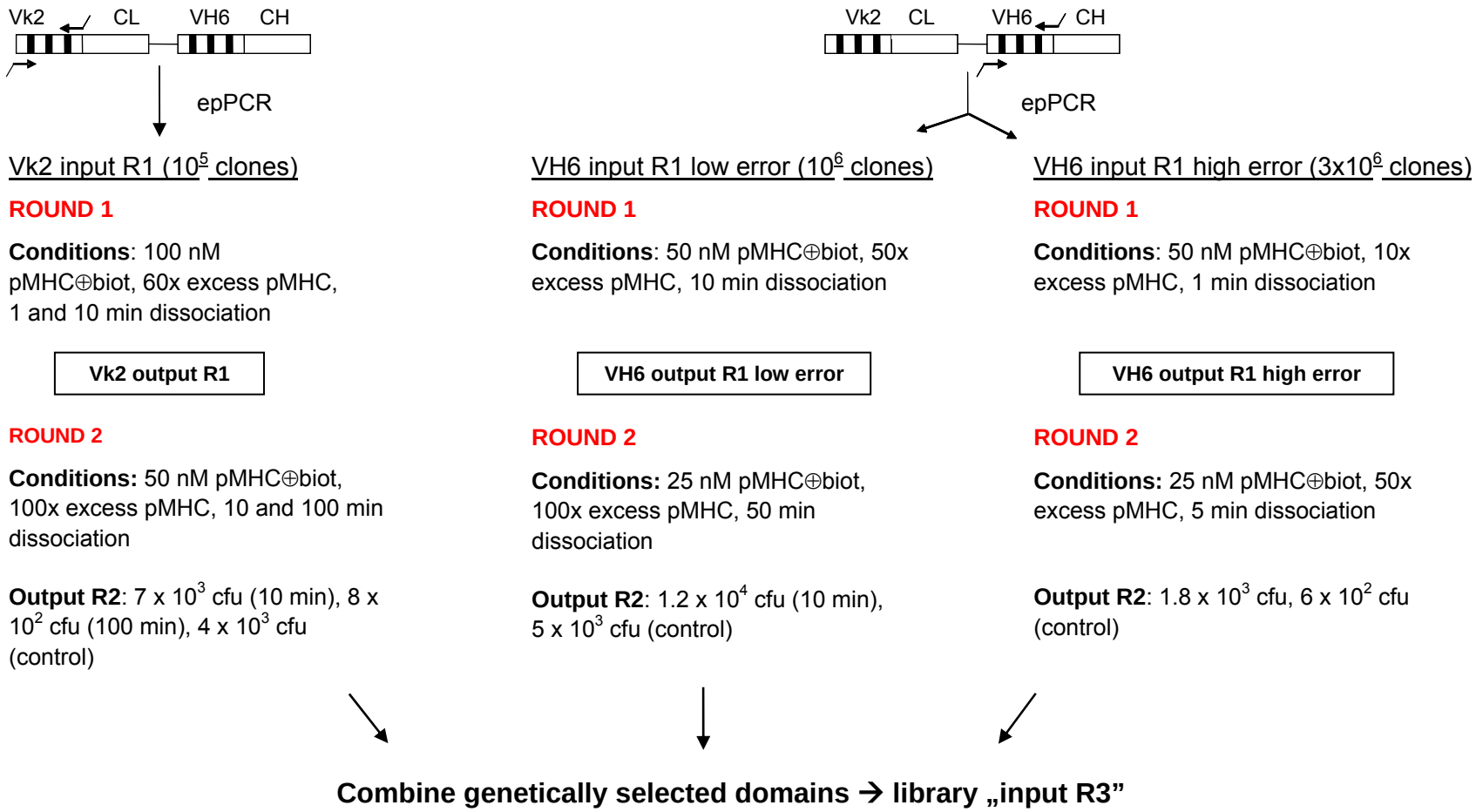


Figure 3-3. Selection scheme for off-rate selection. On the top, the black bars symbolize the CDRs. The output observed after every selection round represents the number of phages recovered for each selection (cfu) compared to a selection control where the biotinylated antigen is absent.

Single clone analysis of members of the “input R3” library was performed in ELISA format: 92 clones were expressed as Fab fragments in microtiter plates, and cells were lysed chemically by incubating the culture with lysozyme and tested for binding to MHC-gp33, MHC-ASIA and β2-microglobulin, all immobilized onto Maxisorp plates. A dozen clones showed higher signals (higher absorption at 405 nm), compared to ConFab and were further tested for specificity against unrelated peptide-MHC complexes (Figure 3-4). These promising clones were sequenced. Curiously, the wild-type parental variable domain of the heavy chain has been re-selected for all of them. All the selected Fabs contained the same heavy chain sequence as the ConFab, on the amino acid and on the nucleotide level (data not shown). All the mutations have been accumulated in the variable domain of the light chain. They were evenly distributed between the CDRs and the framework regions with the notable exception that the first framework region did not contain any mutated residue. Moreover, it was observed that tyrosine residues were quite frequently mutated to cysteine. An alignement of a subset of the selected Fabs is shown in Figure 3-5.

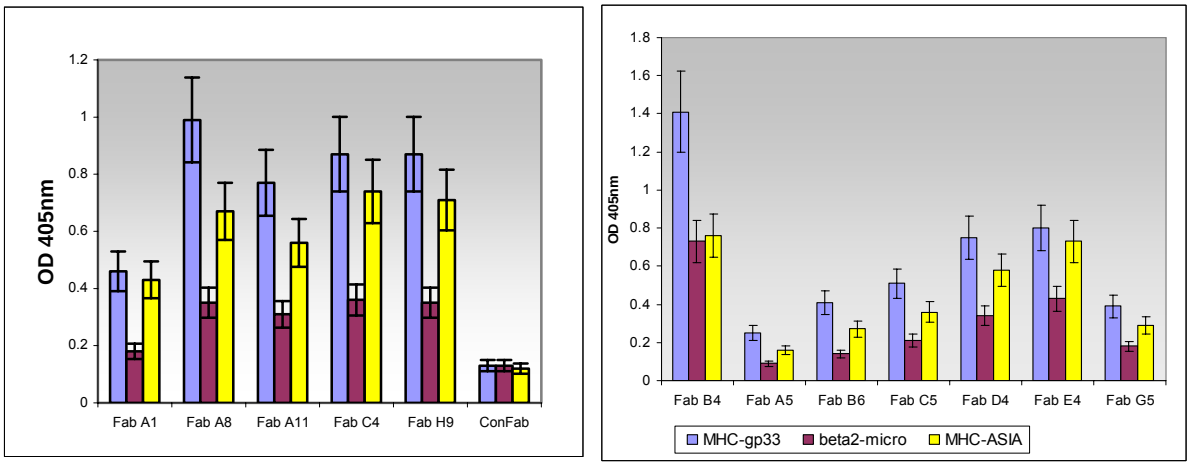


Figure 3-4. Crude extract ELISA performed with single clones of the “input R3” library. After a first initial screening, a dozen Fabs were further investigated with regard to their binding specificity against MHC-gp33 (blue), MHC-ASIA (yellow) and β2-microglobulin (purple) coated onto a Maxisorp plate. Fabs bound to these antigens were detected with an anti-Flag M2 and goat anti-mouse IgG coupled to alkaline phosphatase. The background has been subtracted (well containing the coated antigen, the detection antibodies, but no Fab fragment).

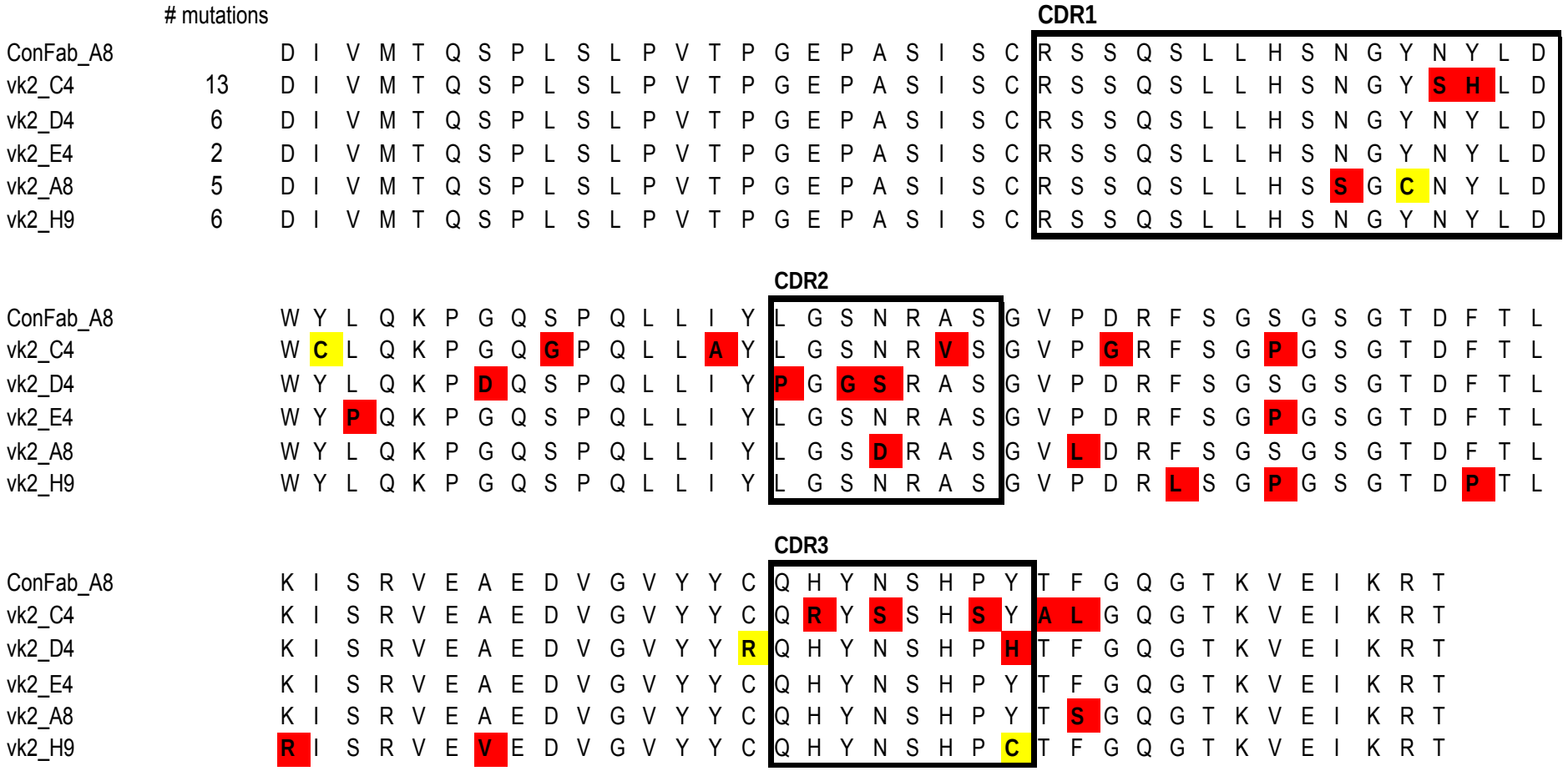


Figure 3-5. Alignment of selected library members from the "Vk2 low error" library. Each clone is denominated by a letter and a number. The second column indicates the number of mutations for each clone. Note that no mutations affect the framework region located before the CDR1. Every clone (except E4) contains an additional or a removed cysteine residue (highlighted in yellow).

Aliquots of an overnight expression culture were loaded on an SDS-PAGE and blotted separately for detection via the anti-TetraHis antibody (detecting the heavy chain) or the human anti-kappa light chain (binding the Vκ2 domain). Surprisingly, the light chain of every selected Fab was found in the insoluble fraction (Figure 3-6). Presence of an additional cysteine residue or removal of an existing one could not be the sole reason for this observation as Fab_E4, for instance, also showed a precipitated light chain, whereas none of its mutations affected cysteine residues (see Figure 3-5). In the Western blot detecting the heavy chain, a very faint band is present in the soluble fraction of Fab_A8, Fab_H9 and of even stronger intensity for Fab_E4. These three clones were expressed on a bigger scale and the Fabs were purified through a Nickel-NTA affinity column (IMAC). By subjecting the purified proteins to a mass spectrometry analysis, it turned out that we purified the complete heavy chain of the Fab (variable and constant domains) in complex with the cytoplasmic chaperone GroEL. This association might likely have arisen during bacterial lysis previous to the purification procedure. As expected, addition of adenosine triphosphate in order to dissociate the chaperone resulted in the precipitation of the heavy chain of the antibody (data not shown).

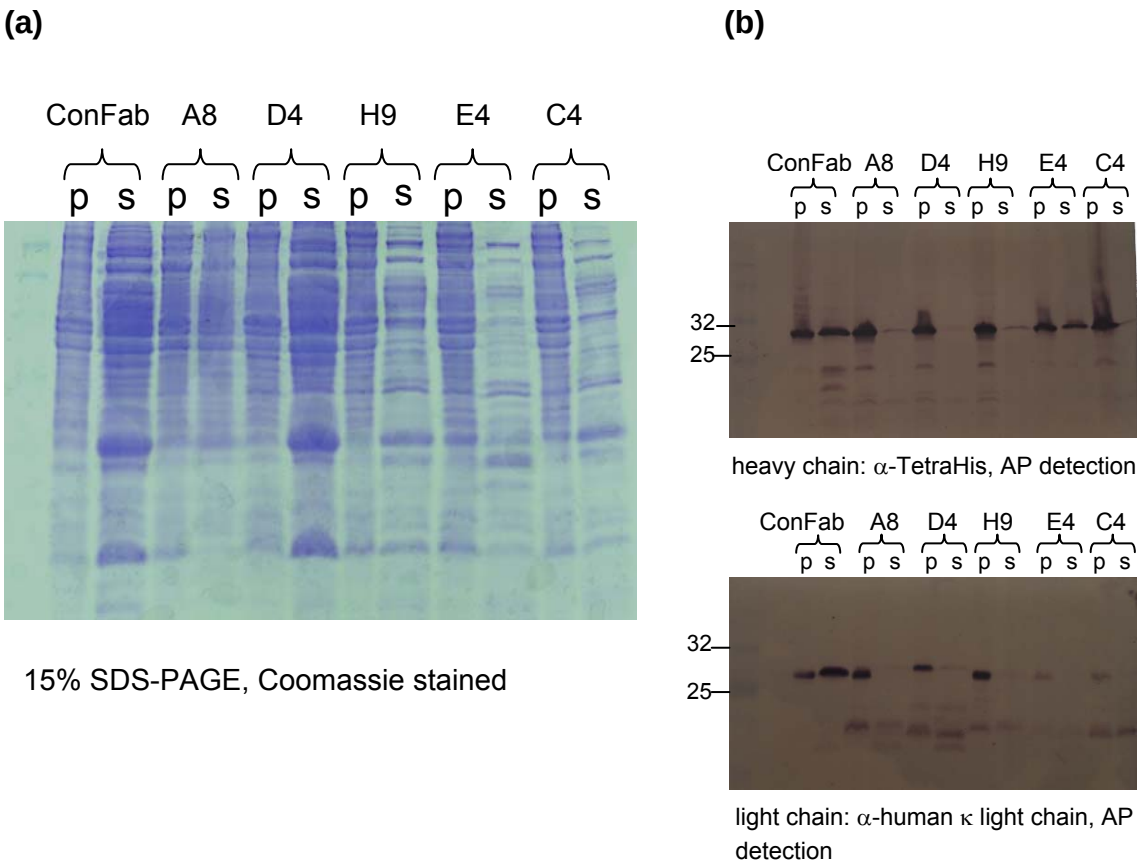
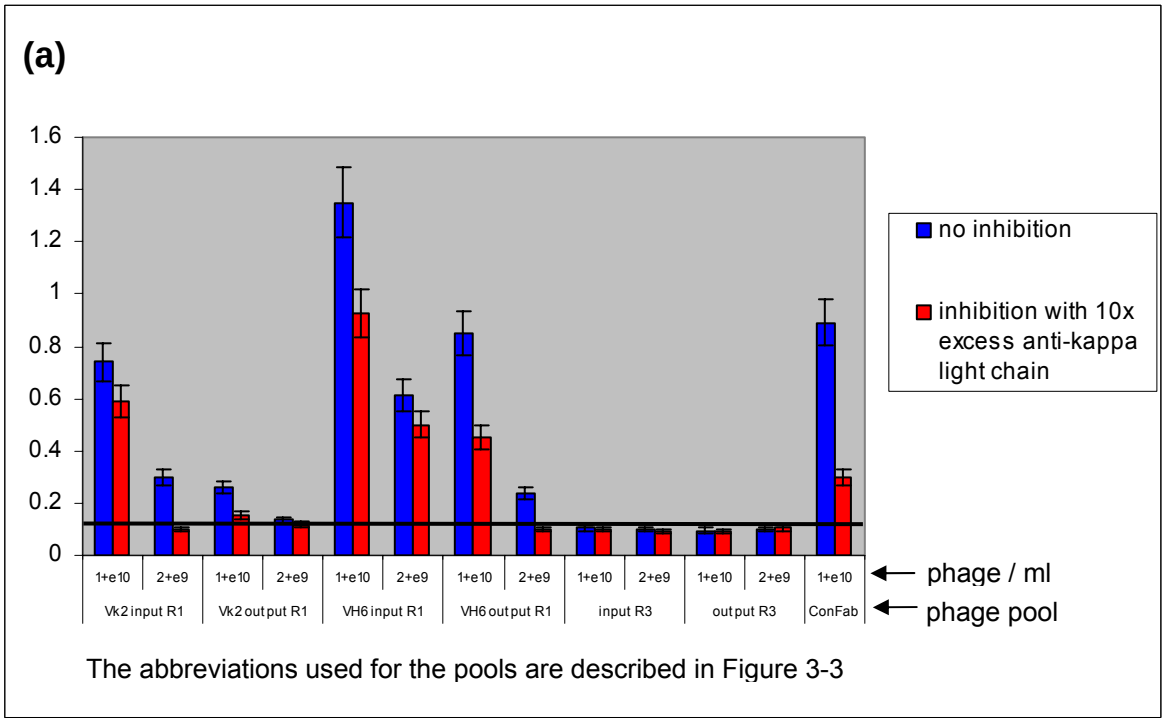


Figure 3-6. (a) The Fab fragments have been periplasmically expressed at 22°C. After overnight growth, cells have been centrifuged, resuspended, lysed by French Press and centrifuged again to separate the soluble (s) and the insoluble fraction (p). (b) On a Western blot, the light chains are mostly present in the insoluble fraction, whereas a tiny fraction of the heavy chain of Fab_A8, Fab_H9 and Fab_E4 remains in the soluble fraction. The numbers on the left indicate molecular weight standards (in kilodaltons). ConFab stands for the original wild-type Fab fragment used for randomization.

In view of this quite unexpected result, we decided to analyze the phage pools of each library, collected after each selection round. The binding of the phage to MHC-gp33 was investigated, as well as the binding to an anti-human kappa light chain antibody. The rationale behind this experiment was to investigate whether the selected phage were still able to recognize the peptide-MHC complex and display the light chain at their tip, which is only associated when the heavy chain is fused to pIII. The two phage ELISAs shown in Figure 3-7 brought some informative insights into the mechanism found by the system to escape the selection pressure. One can immediately notice the total absence of binding to MHC-gp33 for all of the pools, with the notable exception of the “VH6 output R1” (output of the first selection round of the VH6 library). The binding signal unequivocally lies below the background threshold for most of the analyzed pools (Figure 3-7, upper panel). This was not unexpected since the randomization procedure performed by error-prone PCR will have an impact on the integrity of the library members, and consequently lead to non-functionality for a fraction of them. This is most clearly seen for the light chain, whose presence was investigated with an anti-kappa antibody: after one single selection round on the Vκ2 library, the signal for binding to the coated anti-kappa antibody was dramatically reduced (Figure 3-7, upper panel; see “Vκ2 output R1”). However, surprisingly, none of the phage in the pool of the “input R3” library displayed a Fab with a functional light chain able to bind the anti-kappa light chain antibody coated onto the ELISA plate (Figure 3-7b). One has to stress that these ELISAs were performed with pools of phage, and by no way did it tell us anything about the presence or absence of the light chain for single clones. As shown below, at least one clone still displayed its light chain at the phage tip but its signal was quenched in this phage ELISA by the overwhelming majority of the clones which had lost it.

Considering the outcome of this experiment, we investigated the pool of the VH6 library after one selection round in greater detail (“VH6 output R1”, in Figure 3-7). This pool seemed to be slightly enriched in potential binders to MHC-gp33 and the light chain still retained its ability to bind the coated anti-kappa light chain antibody. However, it turned out that the original ConFab had been re-selected, attesting to the fact that the selection had converged to a single and unique sequence. Thus, after two selection rounds, the original nucleotide sequence of the parental ConFab had been recovered for all members of the VH6 library, no matter if the clones were originating from the “high error” or the “low error “ sublibrary (data not shown). Considering these disappointing results, the clone Fab_A8 from the “input R3 library” was further investigated, as this clone seemed to be the most promising one in ELISA performed with crude extracts of proteins (Figure 3-4). The DNA encoding the nucleotide sequence of the selected Fab_A8 was subcloned into the vector pMorph18 in order to be displayed at the tip of the phage particle. Binding and inhibition to various coated antigens were tested.



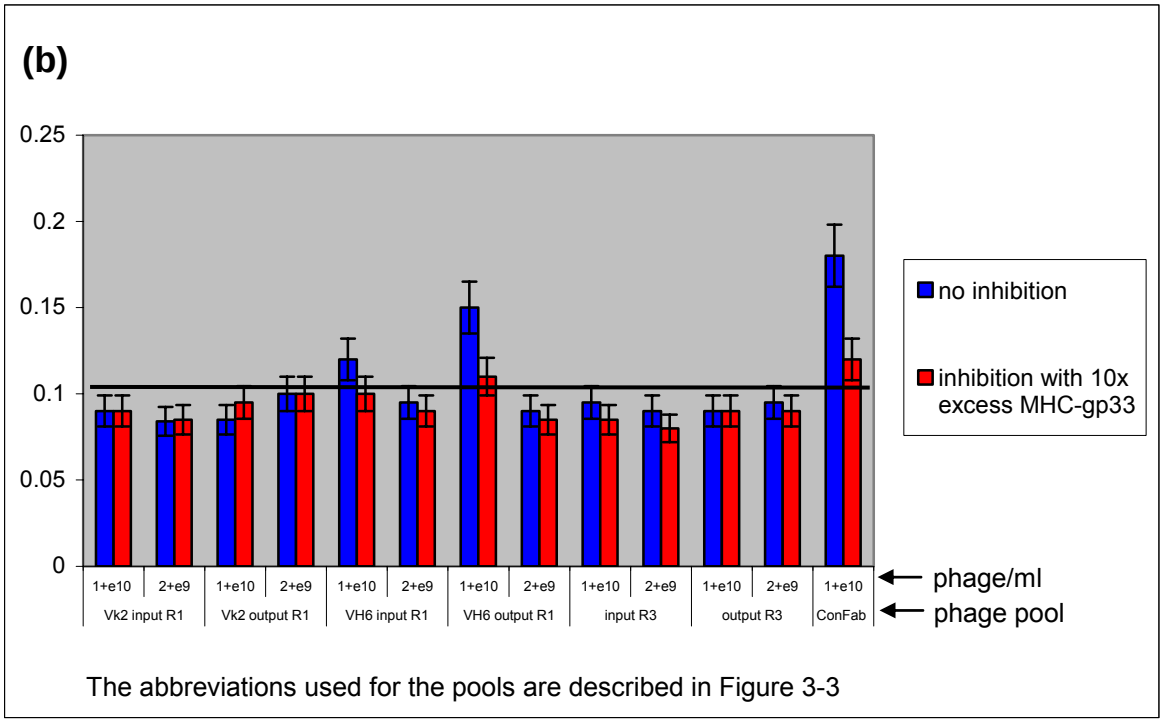


Figure 3-7. Phage ELISAs: analysis of selected pools for binding either to anti-kappa light chain antibody (a) or MHC-gp33 (b). Phage pools after the first and before the third selection round were incubated with immobilized antigen, either with anti-kappa light chain antibody (a) or MHC-gp33 (b) at indicated dilutions ($1x10^{10}$ or $2x10^{9}$ phage/ml). The ConFab (parental clone) is included as a control on the far right side of each plot. For inhibition experiments, phage pools were pre-incubated with excess antigen for 1 hour at RT under constant agitation. Note the different scale on the y-axis in (a) and (b). The horizontal line symbolizes the background level, measured from a well containing the same reagents, except the phage. The measurements were carried out in triplicates.

Very surprisingly, the selected Fab_A8 gave a strong signal for binding to the coated anti-kappa light chain which could be inhibited with soluble anti-kappa antibody, thus demonstrating the presence of a functional light chain displayed at the tip of the phage (Figure 3-8, right). A binding signal to MHC-gp33 could be detected as well. Although the signal was very weak after background subtraction, the binding to MHC-gp33 could be inhibited with an excess of MHC-gp33 in solution, giving credit to the idea that this clone might be specific (Figure 3-8, left). Moreover, it could still discriminate between the gp33 and ASIA peptide, in full accord with the conclusions drawn from the ELISA performed with the periplasmically-expressed protein (Figure 3-4): the signal intensity between the two MHCs differed by around 50% in both ELISA formats. This type of assay has been conducted in the same way with the ConFab displayed on phage; however the signal was too weak to obtain any reliable results (data not shown).

Taking into account the fact that the selected Fab_A8 lost its light chain upon expression in the periplasm of *E. coli* (Figure 3-6) but was still retaining binding to the MHC-gp33 (Figure 3-4), we investigated whether the single heavy chain might still display some residual binding and specificity. The entire variable domain of the heavy chain (VH6 subtype) was PCR amplified and inserted into the vector pTFT74, thus allowing expression in inclusion bodies. As suggested elsewhere, such domains are heavily prone to aggregation when expressed in *E. coli* (Ewert et al., 2003a; Ewert et al., 2003b). This single-domain VH6 antibody ("sdVH6") was expressed in BL21(DE3) strain and refolded at 4°C (see "*Materials & Methods*" for details). The protein had to be prepared freshly for each experiment, as it tended to precipitate after 2 days storage at 4°C.

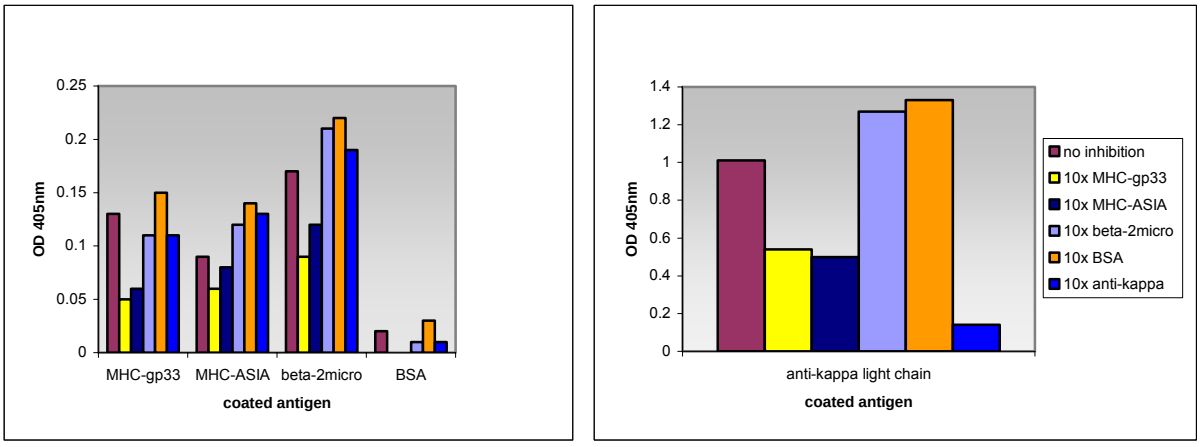
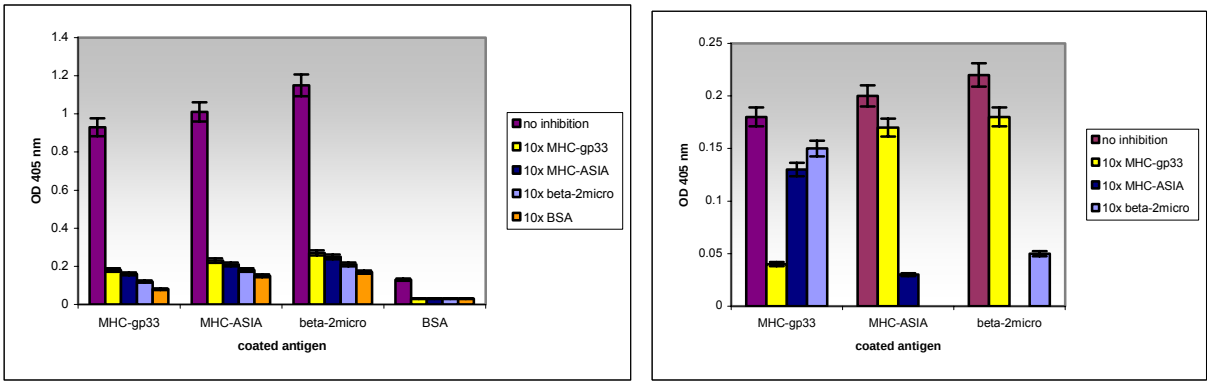


Figure 3-8. Phage ELISAs of the selected Fab_A8, isolated from the input R3 library. The antigens indicated MHC-gp33, MHC-ASIA, BSA (left plot) and anti-kappa light chain antibody (right plot) were coated on a Maxisorp plate. An equal phage concentration (10^{10} phage/ml) was pipetted into each well. For inhibition experiments, phage were pre-incubated with excess antigen (indicated by color codes) or buffer (for the "no inhibition" sample) during 1 hour at RT under constant agitation, before incubation with the coated antigen. Detection was performed with an anti-M13 antibody coupled to horseradish peroxidase (see *Materials & Methods").* Note the different scale on the y-axis in the left and right plot. Background was subtracted similarly as done in Figure 3-7.

The day after purification, this protein preparation was used in an ELISA assay in order to test for binding to MHC-gp33. The results depicted in Figure 3-9 confirmed our hypothesis: the single-domain VH6 antibody was able to bind to MHC-gp33, and this reaction could be inhibited if ten times excess of free soluble MHC-gp33 was present in the well. This reaction showed a reasonable degree of specificity, as a much smaller signal was detected if BSA was coated onto the plate. However, the sdVH6 was not able to discriminate between the gp33 and the ASIA peptide in the context of an MHC molecule. Furthermore, its binding to MHC-gp33 could also be inhibited by excess of BSA. For comparison, the same experiment has been performed with the parental ConFab, and the binding pattern looked much the same (Figure 3-9). It could not distinguish between the gp33 and the ASIA peptide as well, and it sticked to the same extent to coated β2-microglobulin. The amount of protein (5 μM) used in this assay in order to detect a reliable signal hinted at a fairly poor binding constant, thus precluding any visualization of a complex formation on a gel filtration column. The same experiment performed with 1 μM protein gave a signal at the background level for both the single-domain VH6 and the ConFab (data not shown). One could immediately notice, however, that the binding signal to MHC-gp33 was clearly higher (around 5-fold) for the single-domain VH6 antibody, which let us suppose that the affinity may have been slightly improved.

(a)



(b)

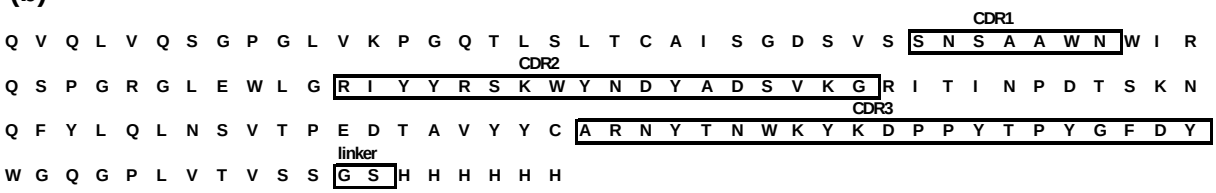


Figure 3-9. ELISAs with the purified single-domain VH6 antibody (left plot) and ConFab (right plot). (a) In both cases, 5 μM purified protein was used. The signal for binding to the MHC-gp33 is significantly higher for the single-domain VH6 (OD_{405nm} = 0.9) than for the parental ConFab (OD_{405nm} = 0.18). Overall both antibodies behave pretty much the same with regard to their binding to other antigens. Background has been subtracted. (b) Amino acid sequence of the single-domain VH6 antibody, corresponding to the variable domain of the heavy chain alone (VH6 subtype). The three CDRs are boxed. A two-amino acid linker was inserted between the protein and the C-terminal hexa-histidine tag. This sequence is identical to the sequence of the variable domain of the heavy chain of the parental ConFab.

3.4 Discussion

We isolated from our phage display libraries an antibody performing as well as (or "as poorly as"!) the parental ConFab. The parental ConFab cannot discriminate between the ASIA and gp33 peptide, nor can the single-domain VH6 antibody. Obviously, the off-rate selection did not restore the peptide specificity. Furthermore, both antibodies stick to coated β2-microglobulin for unknown reasons (Figure 3-9), although a large molar excess of this antigen was added as a competitor during all panning steps. In ELISA performed with purified proteins, both antibodies could be inhibited with an excess of MHC-gp33 in solution. As for the affinity, we cannot affirm to have significantly improved the binding constant, if one takes into consideration the ELISA signals in Figure 3-9. The high concentration (5 μM) of proteins needed to detect a reliable signal hints at a fairly low affinity, which precludes any more precise measurement (e.g. BIACORE). Ideally, the off-rate selection should have resulted in a clone showing a ten times improved binding constant compared to the ConFab Therefore, it is a fair statement to say that we did not greatly improve the binding characteristics of the ConFab, but we also did not make them worse. We eventually fished out of the libraries a shortened version of the parental clone, trimmed down to its minimal component, sufficient for binding the targeted antigen.

It is remarkable that none of the selected Fabs contained a randomized VH domain. The sequences perfectly matched those of the parental ConFab, on the amino acid as well as on the nucleotide level. The reason for this phenomenon is unknown, but one notices that a clone containing a VH domain with a nucleotide sequence exactly identical to the ConFab already existed in the input library, VH6 input R1 high error (data not shown). This clone might have profited from a selective advantage and predominated over the other library members during the selections.

Another issue which is worthwile being considered is the case of the light chain domain, and especially, why the selected Fabs lost their association with the light chain. It is widely accepted that the overall stability of recombinant antibody molecules depends on the intrinsic structural stability of the heavy- and light chain as well as on the extrinsic stabilization provided by their interaction (reviewed in Worn and Plückthun, 2001). We can therefore assume that the intrinsic stability of all the selected light chain domains has been severely affected through the randomization procedure, thereby leading to their aggregation in the periplasm (Figure 3-6b, lower blot). It was expected that a fraction of them would lose their solubility and/or functionality upon library construction, but that all of them were rendered inactive was very surprising. The addition or removal of cysteine residues cannot by itself account for the aggregation tendency of the selected light chains: one selected clone (Fab_E4) has no mutations involving a cysteine residue (Figure 3-5), yet its periplasmic expression does not yield a soluble light chain either. Precipitation of the light chain domains resulted in heavy chain domains having no partner to pair with in the periplasm. This had the major consequence that most of the heavy chains also precipitated, due to hydrophobic patches exposure of the former heavy- and light chain interface. However, when lysing the bacteria, a proportion of heavy chain fragment remained tightly bound to the cytoplasmic chaperone GroEL, thereby ensuring a tiny fraction of correctly folded and functional heavy chain domains. This was especially the case for Fab_A8,

Fab_H9 and to a greater extent for Fab_E4 (Figure 3-6b, upper blot). Furthermore, if one postulates, as suggested elsewhere (Hülsmeyer et al., 2005), that most of the direct contacts between the selected Fabs and the MHC-gp33 peptide complex are mediated by the heavy chain, an antibody consisting on a single heavy chain could still fulfil its function. The absence of the light chain certainly had a severe impact on the overall stability of the heavy chain: the hydrophobic interface became thereby much more exposed to the solvent. But the GroEL chaperone could enhance the folding of the single heavy chain fragment and protects its exposed hydrophobic interface. This complex formation obviously did not affect binding, as the gp33 and ASIA peptides could even be discriminated in the MHC context (Figure 3-4). The experiments described later on with the engineered single domain VH6 antibody became possible, maybe due to the unusually 21 amino acid long CDR3 of VH. It has been suggested that a long CDR-H3 may partially cover the hydrophobic interface region to VL (Ewert et al., 2003b).

Finally, one fundamental question remains to be asked: did the system do what we wanted it to do? In other words, does the selected Fab_A8 displayed at the phage tip show improved binding properties compared to the ConFab? Having a closer look at the Figure 3-8, one can answer yes. In the phage ELISA format, the selected Fab_A8 does show a slightly preferential binding to MHC-gp33 (as opposed to the MHC-ASIA) and is still composed of a heavy chain non-covalently associated with a light chain. The weakness of the ELISA signal (especially for the ConFab, data not shown) precludes any suppositions regarding the magnitude of the affinity improvement. If it did really occur, it certainly does not lie in the desired range. In the Fab ELISA format, the 5-fold higher signal of the sdVH6 bound to MHC-gp33 (Figure 3-9) is appreciable, but the high concentrations of purified proteins used in this assay prevent us from drawing reliable conclusions about the affinity increase. It is, however, remarkable that the selected Fab_A8 retained its full integrity when displayed at the phage tip, meaning that the light chain was still non-covalently associated with the heavy chain when fused to pIII and expressed in the phagemid format. The loss of the light chain must then be due to a problem occurring during the periplasmic over-expression of the Fab in the expression vector. The reason for this different behavior of the selected Fab_A8 in the two vector systems is unknown.

The isolation of a single-domain antibody after panning against an antibody library is not uncommon. It has been recently reported that a single light-chain antibody (derived from an affinity-matured scFv) was able to prevent aggregation of huntingtin. The authors claimed that this is the first reported case where an scFv paratope was mapped entirely to the light chain domain (Colby et al., 2004). On the contrary, a recent publication shows the preponderant role played by the heavy chain for the recognition of a peptide-MHC complex (Figure 3-10).

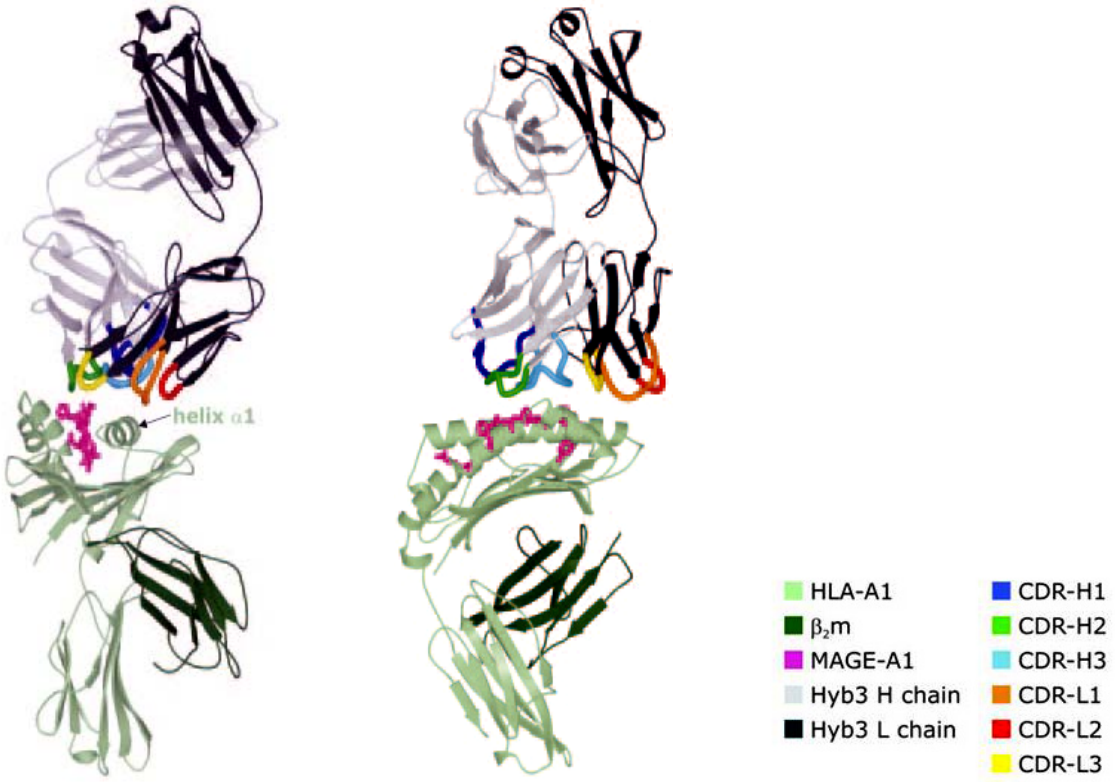


Figure 3-10. Overall topography of the "Hyb3" antibody in complex with the HLA-A1:MAGE-A1 complex. The MAGE-A1 peptide is colored in pink; its C-terminus is in front (left), or rotated by 90° (right). The color code for the Fab, pMHC and CDR loops is given in the legend (Hülsmeyer et al., 2005).

The affinity-matured Fab fragment "Hyb3" contacts the HLA-A1:MAGE-A1 peptide-MHC complex mainly with its CDR-H3, whereas the CDR-L1 and L2 contribute only marginally. In the light chain, only the CDR-L3 makes a direct contribution to the interaction by contacting the α-helix of the MHC (Hülsmeyer et al., 2005). This finding clearly illustrates that the heavy chain mediates most of the direct contacts with the antigen, adding credence to the proposal that an antibody consisting of a single heavy chain might still fulfil its function. The affinity of the starting clone (250 μM) could be markedly improved; the best improved variant showed an affinity of 14 nM as directly measured by BIACORE. This order of magnitude in the binding improvement has already been achieved for many different targets, and since the first paper about affinity maturation published in 1992 (Hawkins et al., 1992), much higher affinity improvements could be realised (Boder et al., 2000; Holler et al., 2000; Jermutus et al., 2001; Low et al., 1996; Schier et al., 1996; Zahnd et al., 2004).

There exist alternative hypotheses to rationalize why the light chain gets so easily eliminated during directed evolution. An attractive idea can be found in recent protein evolution studies dealing with the plasticity of proteins and their resistance to amino acid mutations. Arnold and coworkers postulate that the fractions of folded and thus functional mutants gained from a directed evolution experiment can be significantly increased if thermodynamically stable parents are used. They provide an algorithm for predicting

which structures are better amenable to a high mutational load (Bloom et al., 2005). In our particular case, it could be interesting to check if the low tolerance for mutations of the ConFab is due to a poor thermodynamic stability. This opinion is counterbalanced by other authors who notice that sequence plasticity may be an intrinsic feature of naturally occurring proteins, giving credence to the widespread belief that many experimental mutations have a significant probability in resulting in unchanged or increased stability, but sometimes at the expense of functionality (Taverna and Goldstein, 2002). This dilemma between stability and functionality is probably what the system had to cope with in our selection, and at a certain time it had to choose between the two options in order to circumvent the selection pressure. Obviously, it decided to preserve functionality. And since we selected for binding, it did what we wanted it to do.

3.5 Literature

Aharoni, R., Teitelbaum, D., Arnon, R., and Puri, J. (1991). Immunomodulation of experimental allergic encephalomyelitis by antibodies to the antigen-Ia complex. Nature 351, 147-150.

Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C., and Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci USA 102, 606-611.

Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci USA 97, 10701-10705.

Chames, P., Hufton, S. E., Coulie, P. G., Uchanska-Ziegler, B., and Hoogenboom, H. R. (2000). Direct selection of a human antibody fragment directed against the tumor T-cell epitope HLA-A1-MAGE-A1 from a nonimmunized phage-Fab library. Proc Natl Acad Sci USA 97, 7969-7974.

Colby, D. W., Garg, P., Holden, T., Chao, G., Webster, J. M., Messer, A., Ingram, V. M., and Wittrup, K. D. (2004). Development of a human light chain variable domain (V(L)) intracellular antibody specific for the amino terminus of huntingtin via yeast surface display. J Mol Biol 342, 901-912.

Day, P. M., Yewdell, J. W., Porgador, A., Germain, R. N., and Bennink, J. R. (1997). Direct delivery of exogenous MHC class I molecule-binding oligopeptides to the endoplasmic reticulum of viable cells. Proc Natl Acad Sci USA 94, 8064-8069.

Denkberg, G., Cohen, C. J., Lev, A., Chames, P., Hoogenboom, H. R., and Reiter, Y. (2002). Direct visualization of distinct T cell epitopes derived from a melanoma tumor-associated antigen by using human recombinant antibodies with MHC- restricted T cell receptor-like specificity. Proc Natl Acad Sci USA 99, 9421-9426.

Ewert, S., Honegger, A., and Plückthun, A. (2003a). Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. Biochemistry 42, 1517-1528.

Ewert, S., Huber, T., Honegger, A., and Plückthun, A. (2003b). Biophysical properties of human antibody variable domains. J Mol Biol 325, 531-553.

Forrer, P., and Jaussi, R. (1998). High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage lambda head protein D. Gene 224, 45-52.

Hawkins, R. E., Russell, S. J., and Winter, G. (1992). Selection of phage antibodies by binding affinity. Mimicking affinity maturation. J Mol Biol 226, 889-896.

Hennecke, J., and Wiley, D. C. (2001). T cell receptor-MHC interactions up close. Cell 104, 1-4.

Holler, P. D., Chlewicki, L. K., and Kranz, D. M. (2003). TCRs with high affinity for foreign pMHC show self-reactivity. Nat Immunol 4, 55-62.

Holler, P. D., Holman, P. O., Shusta, E. V., O'Herrin, S., Wittrup, K. D., and Kranz, D. M. (2000). In vitro evolution of a T cell receptor with high affinity for peptide/MHC. Proc Natl Acad Sci USA 97, 5387-5392.

Housset, D., and Malissen, B. (2003). What do TCR-pMHC crystal structures teach us about MHC restriction and alloreactivity? Trends Immunol 24, 429-437.

Hülsmeyer, M., Chames, P., Hillig, R. C., Stanfield, R. L., Held, G., Coulie, P. G., Alings, C., Wille, G., Saenger, W., Uchanska-Ziegler, B., et al. (2005). A major histocompatibility complex-peptide-restricted antibody and T cell receptor molecules recognize their target by distinct binding modes: crystal structure of human leukocyte antigen (HLA)-A1-MAGE-A1 in complex with FAB-HYB3. J Biol Chem 280, 2972-2980.

Inaba, K., Pack, M., Inaba, M., Sakuta, H., Isdell, F., and Steinman, R. M. (1997). High levels of a major histocompatibility complex II-self peptide complex on dendritic cells from the T cell areas of lymph nodes. J Exp Med 186, 665-672.

Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., and Plückthun, A. (2001). Tailoring in vitro evolution for protein affinity or stability. Proc Natl Acad Sci USA 98, 75-80.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Plückthun, A., and Virnekas, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol 296, 57-86.

Lev, A., Denkberg, G., Cohen, C. J., Tzukerman, M., Skorecki, K. L., Chames, P., Hoogenboom, H. R., and Reiter, Y. (2002). Isolation and characterization of human recombinant antibodies endowed with the antigen-specific, major histocompatibility complex-restricted specificity of T cells directed toward the widely expressed tumor T-cell epitopes of the telomerase catalytic subunit. Cancer Res 62, 3184-3194.

Low, N. M., Holliger, P. H., and Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. J Mol Biol 260, 359-368.

McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature 348, 552-554.

Niedermann, G. (2002). Immunological functions of the proteasome. Curr Top Microbiol Immunol 268, 91-136.

Plaksin, D., Polakova, K., McPhie, P., and Margulies, D. H. (1997). A three-domain T cell receptor is biologically active and specifically stains cell surface MHC/peptide complexes. J Immunol 158, 2218-2227.

Porgador, A., Yewdell, J. W., Deng, Y., Bennink, J. R., and Germain, R. N. (1997). Localization, quantitation, and in situ detection of specific peptide-MHC class I complexes using a monoclonal antibody. Immunity 6, 715-726.

Princiotta, M. F., Finzi, D., Qian, S. B., Gibbs, J., Schuchmann, S., Buttgereit, F., Bennink, J. R., and Yewdell, J. W. (2003). Quantitating protein synthesis, degradation, and endogenous antigen processing. Immunity 18, 343-354.

Puri, J., Arnon, R., Gurevich, E., and Teitelbaum, D. (1997). Modulation of the immune response in multiple sclerosis: production of monoclonal antibodies specific to HLA/myelin basic protein. J Immunol 158, 2471-2476.

Reiter, Y., Di Carlo, A., Fugger, L., Engberg, J., and Pastan, I. (1997). Peptide-specific killing of antigen-presenting cells by a recombinant antibody-toxin fusion protein targeted to major histocompatibility complex/peptide class I complexes with T cell receptor-like specificity. Proc Natl Acad Sci USA 94, 4631-4636.

Saric, T., Chang, S. C., Hattori, A., York, I. A., Markant, S., Rock, K. L., Tsujimoto, M., and Goldberg, A. L. (2002). An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. Nat Immunol 3, 1169-1176.

Schier, R., McCall, A., Adams, G. P., Marshall, K. W., Merritt, H., Yim, M., Crawford, R. S., Weiner, L. M., Marks, C., and Marks, J. D. (1996). Isolation of picomolar affinity anti-c-erbB-2 single-chain Fv by molecular evolution of the complementarity determining regions in the center of the antibody binding site. J Mol Biol 263, 551-567.

Taverna, D. M., and Goldstein, R. A. (2002). Why are proteins so robust to site mutations? J Mol Biol 315, 479-484.

Valitutti, S., Muller, S., Cella, M., Padovan, E., and Lanzavecchia, A. (1995). Serial triggering of many T-cell receptors by a few peptide-MHC complexes. Nature 375, 148-151.

Willcox, B. E., Gao, G. F., Wyer, J. R., Ladbury, J. E., Bell, J. I., Jakobsen, B. K., and van der Merwe, P. A. (1999). TCR binding to peptide-MHC stabilizes a flexible recognition interface. Immunity 10, 357-365.

Worn, A., and Plückthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. J Mol Biol 305, 989-1010.

Wülfing, C., and Plückthun, A. (1994). Correctly folded T-cell receptor fragments in the periplasm of Escherichia coli. Influence of folding catalysts. J Mol Biol 242, 655-669.

Yewdell, J. W. (2001). Not such a dismal science: the economics of protein synthesis, folding, degradation and antigen processing. Trends Cell Biol 11, 294-297.

Yewdell, J. W., and Bennink, J. R. (2001). Cut and trim: generating MHC class I peptide ligands. Curr Opin Immunol 13, 13-18.

York, I. A., Chang, S. C., Saric, T., Keys, J. A., Favreau, J. M., Goldberg, A. L., and Rock, K. L. (2002). The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. Nat Immunol 3, 1177-1184.

Young, A. C., Zhang, W., Sacchettini, J. C., and Nathenson, S. G. (1994). The three-dimensional structure of H-2Db at 2.4 A resolution: implications for antigen-determinant selection. Cell 76, 39-50.

Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J Mol Biol 255, 589-603.

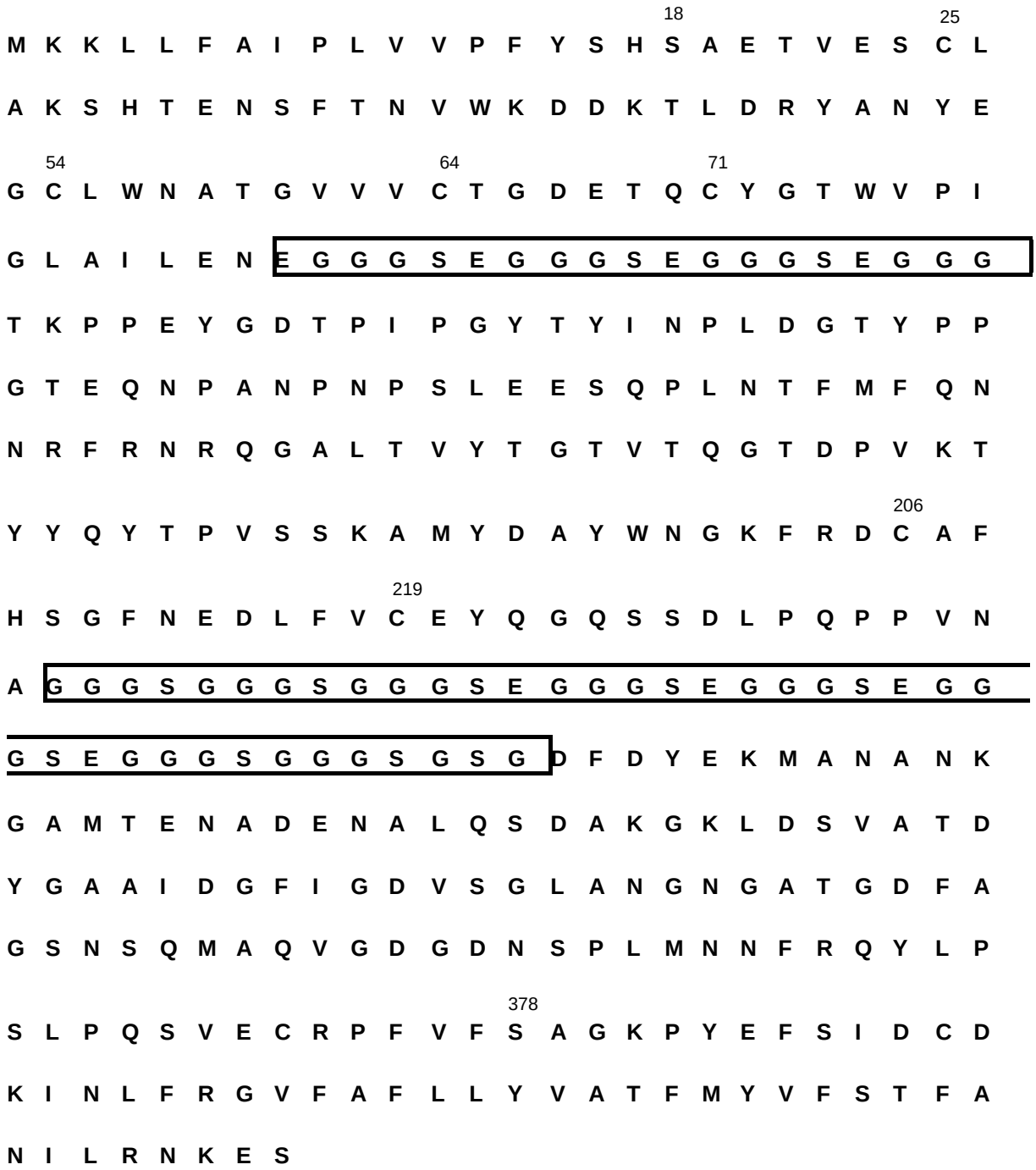
Zahnd, C., Spinelli, S., Lüginbuhl, B., Amstutz, P., Cambillau, C., and Plückthun, A. (2004). Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. J Biol Chem 279, 18870-18877.

Zhong, G., Reis e Sousa, C., and Germain, R. N. (1997). Production, specificity, and functionality of monoclonal antibodies to specific peptide-major histocompatibility complex class II complexes formed by processing of exogenous protein. Proc Natl Acad Sci USA 94, 13856-13861.

Abbreviations

BCIP	5-bromo-4-chloro-3-indolyl phosphate p-toluidine salt
BSA	bovine serum albumin
CDR (e.g. CDR-H1)	complementary determining region (e.g. first CDR of the heavy chain)
DMF	dimethylformamide
dNTP	desoxy-nucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetate
ELISA	enzyme linked immunosorbent assay
epPCR	error-prone PCR
Fab	fragment antigen binding (antigen-binding fragment obtained by papain digestion of an immunoglobulin)
HSQC	heteronuclear single quantum correlation
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
MALDI	matrix assisted laser ion desorption
MHC	major histocompatibility complex
MW	molecular weight
NMR	nuclear magnetic resonance
NBT	nitro blue tetrazolium
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PVDF	polyvinylidene fluoride
RT	room temperature
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
VH/VL	variable domain of the heavy chain / light chain
4-nPP	4-nitrophenyl phosphate
HLA	Human Leukocyte Antigen (human MHC)
MAGE	melanoma-associated antigen

pIII minor coat protein (P03662)



Amino acid sequence of the pIII minor coat protein from bacteriophage M13 (SwissProt entry P03662). The first 18 residues contain the signal sequence. The protein therefore starts at position 19 (N1 domain) . Six cysteine residues form three pairs of disulfide bridges in the native protein: Cys 25 with Cys 54, Cys 64 with Cys 71 and Cys 206 with Cys 219. The glycine-rich linkers are boxed. Ser 378 is replaced by a Glycine in our variant. For more details about constructs used for expression and characterization, see section 2-1, page 27.

TOLR_ECOLI (P05829)

20

M A R A R G R G R R D L K S E I N I V P L L D V L L

40

V L L L I F M A T A P I I T Q S V E V D L P D A T E

S Q A V S S N D N P P V I V E V S G I G Q Y T V V V

E K D R L E R L P P E Q V V A E V S S R F K A N P K

T V F L I G G A K D V P Y D E I I K A L N L L H S A

142

G V K S V G L M T Q P I

Amino acid sequence of the TolR protein (SwissProt entry P05829). The first twenty positions are the probable cytoplasmic part. Residues 20 to 40 contain a putative signal anchor for type II membrane protein. The periplasmic portion starts at position 41. See section 2.2 for further details about constructs used for expression and characterization (page 45).

TONB_ECOLI (P02929)

```
M T L D L P R R F P W P T L L S V C I H G A V V A G
                32
L L Y T S V H Q V I E L P A P A Q P I S V T M V T P
A D L E P P Q A V Q P P P E P V V E P E P E P E P I
                                                103
P E P P K E A P V V I E K P K P K P K P K P K P V K
K V Q E Q P K R D V K P V E S R P A S P F E N T A P
                                              155
A R L T S S T A T A A T S K P V T S V A S G P R A L
              164
S R N Q P Q Y P A R A Q A L R I E G Q V K V K F D V
T P D G R V D N V Q I L S A K P A N M F E R E V K N
A M R R W R Y E P G K P G S G I V V N I L F K I N G
T T E I Q
```

Amino acid sequence of the TonB protein (SwissProt entry P02929). The region anchored in the bacterial inner membrane covers the first 32 amino acids. The segment including a proline-rich region is boxed, this part is thought to exist in an extended conformation in the native protein. The visible part of our crystal structure presented in section 2-3a (page 59) starts at position 164, whereas we cloned the fragment starting at position 155. Another structure, done with a fragment covering residue 164 to 239 also crystallized as a dimer (Koedding et al., 2004). A third structure was solved with a construct including residues 103-239; this domain was found to be monomeric with several unexpected structural elements distinct from the previously determined crystal structures (Sean Peacock et al., 2005).

Literature

Koedding, J., Howard, P., Kaufmann, L., Polzer, P., Lustig, A., and Welte, W. (2004). Dimerization of TonB Is Not Essential for Its Binding to the Outer Membrane Siderophore Receptor FhuA of Escherichia coli. J Biol Chem 279, 9978-9986.

Sean Peacock, R., Weljie, A. M., Peter Howard, S., Price, F. D., and Vogel, H. J. (2005). The Solution Structure of the C-terminal Domain of TonB and Interaction Studies with TonB Box Peptides. J Mol Biol 345, 1185-1197.

Curriculum Vitae

Name	Mooser
First names	Alexandre Pierre
Date and place of birth	August 29th 1973 in Fribourg, Switzerland
Nationality	Swiss

Education

1998-2005	Dissertation in the group of Prof. Dr. Andreas Plückthun, Department of Biochemistry, University of Zürich, Switzerland. Thesis title: ” Filamentous phage infection of *E. coli*: mechanistic aspects and applications to *in vitro* protein evolution”
1994-1998	Diplom in Biology, University of Fribourg, Switzerland. Title of the diploma thesis: “Functional analysis of a PI 3-Kinase Homologue in C.elegans” conducted at the Department of Zoology (Prof. Tobler)
Summer 1996	Practical training at the Department of Biochemistry, University of Zürich, Switzerland (Prof. Hübscher)
1989 - 1993	Gymnasium, Maturatypus B, Collège St-Michel, Fribourg